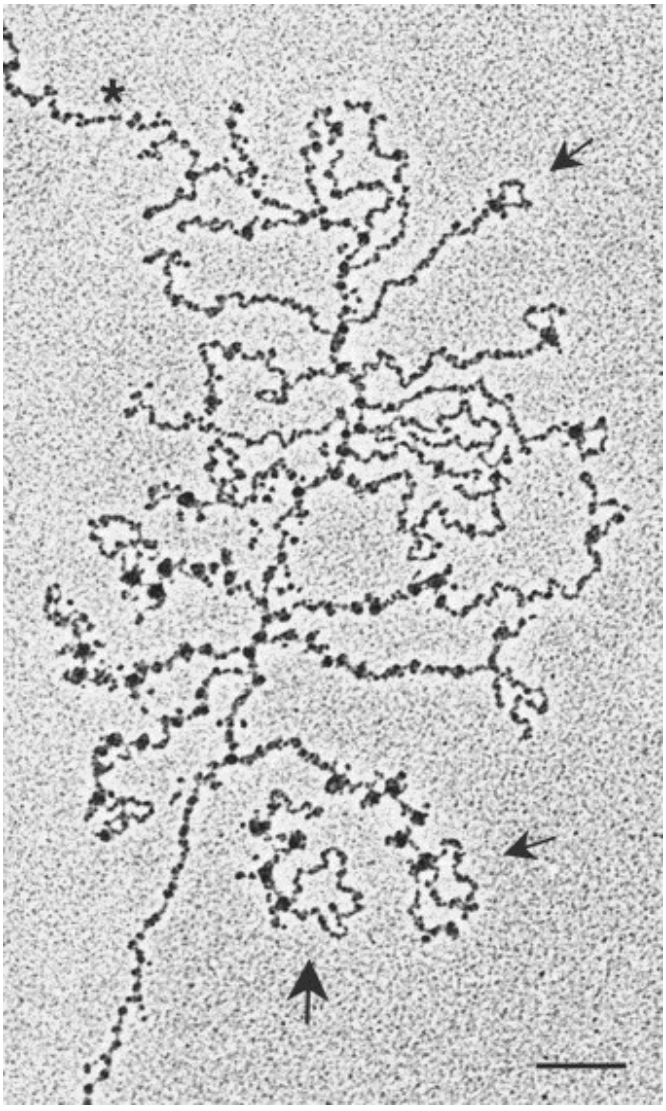
THE UNIVERSITY
of EDINBURGH

Investigations of RNA production and processing in *Saccharomyces cerevisiae*

Thesis presented for the degree of Doctor of Philosophy.
Completed by Research Publications

January 2021

James David Barrass s2018725



Declaration

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated in the Contribution sections of chapters 2 to 7 and the bound publications

David Barrass s2018725 24 January 2021

Front Page: Electron micrograph showing co-transcriptional splicing (splicing occurring as the RNA is being transcribed), in Drosophila. The DNA, with nucleosomes, is the thread running top to bottom. Pre-mRNA transcripts are threads appearing from the DNA, the * is the approximate location of the transcription start site. The spliceosome is the dot near the lariat loops on the RNA (3 arrowed), many RNAs do not have the lariats and could already have undergone splicing. The bottom loop could be transcribed and cleaved RNA. Bar 200 nm.

Image reproduced from Proudfoot, N. (2000). Connecting transcription to messenger RNA processing. Trends Biochem. Sci. 25, 290–293, based on original research originally presented in Beyer AL, Osheim YN (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 2(6):754-65.

Abstract

Over the course of 10 years and 14 publications, 7 of which are described in this thesis, I made significant contributions to the field of RNA processing, particularly RNA splicing in budding yeast. These contributions have been both in knowledge accrued and techniques developed and optimised.

For the Ribo1 reporter (chapter 2), I developed the RT-qPCR assays, RNA copy number/cell estimation and 3' end cleavage assay. These show that the minimum time taken for signal induction, recruitment, assembly of transcription factors and ultimately transcription, is in the order of 4 minutes, with transcription rates of 60 to 90 seconds per kb. The 3' end cleavage assay reveals that the initial pulse of transcripts may be spliced only partially co-transcriptionally; presumably splicing factors are recruited to the site of transcription during this period. After this initial phase, in the reporter at least, splicing is almost exclusively co-transcriptional. Mutant transcripts on which spliceosomes assemble but are unable to complete splicing, are targeted for very rapid degradation.

I have optimised the thio-labelling technique to the point where I can detect thiolated RNAs just 15 seconds after addition of the 4-thiouracil nucleotide analogue to cell cultures (chapter 3). From my data and transcriptome sequencing (chapter 4), I have constructed models of transcription and splicing, which indicate that co-transcriptional splicing is a general feature of most yeast pre-mRNA transcripts and almost all ribosomal protein gene transcripts. Intronic features that act against co-transcriptional splicing include runs of uracils and secondary structure, especially over the branch point adenosine (chapter 4). I have provided a highly detailed protocol and video for thio-labelling of RNA in vivo and its purification (chapter 3).

Similarly, I have developed and prepared a comprehensive protocol for auxin-induced protein depletion, to the point where this is the most flexible and least metabolically perturbing technique for doing this in yeast (chapter 6). Applying this method, whilst thio-labelling, I revealed that depleting splicing factors

sequesters spliceosome components, rapidly resulting in the cessation of splicing (chapter 7). My microarray results indicate that this is a common consequence of splicing factor deactivation for many or all transcripts.

In this thesis I examine RNA metabolism, noting that the processes of transcription, splicing, 3' end formation and degradation are coordinated and harmonised to optimise fidelity, flexibility and efficiency.

Lay Summary

Humans, plants, yeast and bacteria require instructions to make themselves. The instructions are stored on DNA, a very long string-like molecule. Think of this as a length of text approximately 6 billion letters long that has all the instructions to make an organism, in this case you. This amounts to 2 meters of DNA inside every cell of the body. Enough, if a body's worth was joined into one long string, to stretch halfway to the Sun. The instructions are broken down into simpler instructions, called genes, analogous to a sentence or so with details on how to make a small component of the whole organism.

The DNA molecule is however very thin, fragile and prone to tangling, so it must be protected and carefully packed away. So, what does the cell do if it needs to make a component? The cell does what we would do, it makes a copy of the DNA, in this case into a very similar molecule called RNA. This has many advantages: the core instructions on DNA are protected and many copies of the RNA can be made and distributed so more of the component made quickly. Also, the RNA is chemically not as stable as the DNA, so it's easy to remove unwanted instructions when there's no longer a need for that component.

The analogy is: there is a master manual (DNA), held in an office (the cell nucleus). If a new component is needed the relevant instructions are copied down on scraps of paper (RNA), and given to several teams to make lots of that component. Once that task is done the papers are recycled so no one gets

confused and makes the wrong thing. The copying process is called transcription. This, and RNA degradation in a cell, are important in this thesis.

In many organisms there is additional complexity. In the instructions there are often stretches that are not meaningful <start of meaningless text> fkjwklgbwlkg bwgbkjqhvbqwkgbqekjbgkq kjg kbgkjh bgkACUAAUjg g kjgekjbgen <end of meaningless text>. These meaningless sections must be removed, and the two meaningful sections either side spliced together. Only once this is done can the component be made. This splicing process is another important part of this thesis.

Yeast has much, much shorter DNA than a human, but it still must overcome some of the same problems. So, it has simpler systems that are easier to investigate. Lessons about these systems discovered in yeast can be applied to human health. Revealed in this study is that the copying process writes about 1000 characters in a minute. Also, the process of removing the meaningless sections is done whilst the DNA is being copied.

University Regulations on PhD by Research Publications

This thesis has been prepared in accordance to the University of Edinburgh's regulations for PhD by research publications, http://www.drps.ed.ac.uk/20-21/regulations/PGDRPS2020-21.pdf, sections 52 to 55. The work is based on publications to which I have made a significant contribution to in the last ten years (2009 to 2019). A significant contribution is considered to be, at the minimum, being either first or second author of the publication and work leading to a complete figure. Splicing has always been at the heart of the work, but splicing has connections to other processes such as transcription, 3’ end formation and degradation which have to be considered to gain a complete picture of the RNA life cycle.

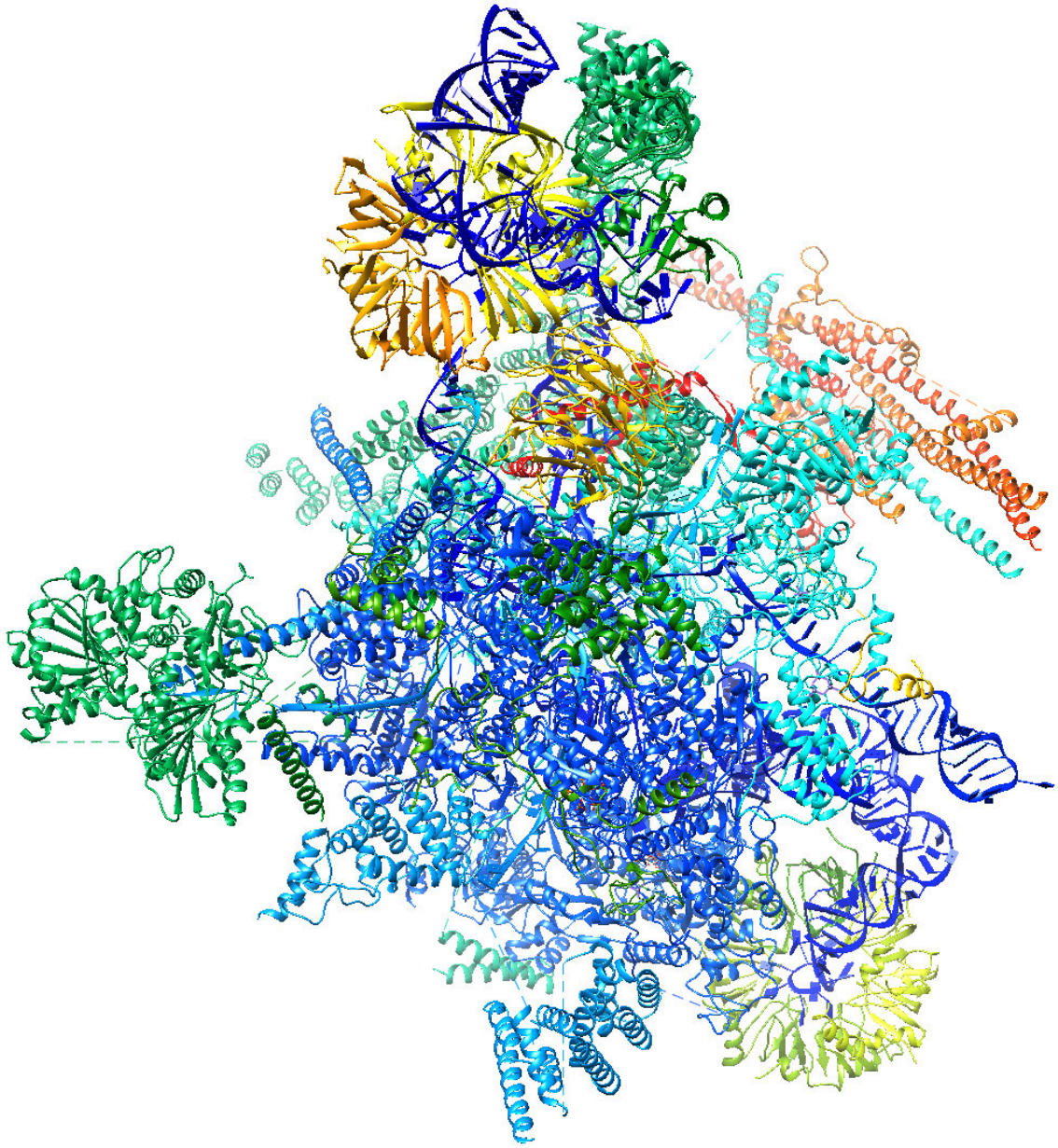
Acknowledgements

It is in this section where one normally gives thanks to parents, I have rather more to thank in my family, but I do want to thank my mother whose belief in me was constant despite my rather scatty nature. I only wish you were still around to see me having got this far. Thanks go, of course, to my wife - Clair who again has always supported me. I really cannot thank you enough especially as I know you have put up with a lot while I've been doing this. Thank you to my children, Stephen for providing competition to see who could get their PhD first. Chrissie for putting up with me and for creating the magazine cover shown at the start of Chapter 2. I have to thank my cat, who has seen me sitting writing for long periods of time as a wonderful opportunity to sleep. I have continued writing much longer than I intended, though I don't thank her for the back ache it's caused.

Thanks to all the Beggs lab members, you have all been wonderful people. I have to thank all my collaborators in each paper, but I'd like to extend special thanks to Ross, Steve, Ema and Susanna for encouraging me to do a PhD. Of course, thanks go to Jean for employing me for almost 20 years and allowing me the space to do these experiments. Your encyclopaedic knowledge of splicing has been a constant source of wonder to me and some very powerful experiments.

Contents

1 Introduction



The spliceosome:
Cryo-electron microscopy structure of the *Saccharomyces cerevisiae* spliceosome at the B* stage coloured by chain (Wan et al., 2019)

The central dogma of molecular biology states that information flows from DNA, through RNA, and into protein (Crick, 1970). DNA provides the genetic information store and protein provides structure and performs catalytic work for the cell. This is mostly right and broadly true, but views RNA as merely an intermediate. RNA is not passive but undergoes and performs many processing events that affect DNA replication and protein synthesis as well as its own stability. This study examines how RNA is prepared for protein production and how this process is made efficient by occurring while the RNA is being transcribed from DNA, i.e. co-transcriptionally (Carrocci and Neugebauer, 2020) .

1.1 Transcription

The process of transcription is the copying of a DNA template into RNA (Svetlov and Nudler, 2013). In the nuclei of eukaryotic cells this is performed by 3 DNA-directed RNA polymerases (plants have 2 additional polymerases); RNA polymerase I (RNAPI), RNA polymerase II (RNAPII) and RNA polymerase III (RNAPIII). RNAPI transcribes the larger rRNAs, but not 5S rRNA. RNAPIII is responsible for tRNAs, 5S rRNA, and some small nuclear and nucleolar RNAs (snRNAs & snoRNAs) including the U6 snRNA involved in splicing. Although RNAPI transcribes only 3 RNAs these are highly abundant and along with RNAPIII contributes approximately 95% of the cell's RNA content.

RNAPII produces pre-messenger RNA (pre-mRNA) transcripts, some small RNAs and longer non-coding RNAs. Of note in the latter category are cryptic unstable transcripts (CUTs) (Wyers et al., 2005), stable unannotated transcripts (SUTs) (Xu et al., 2009), and Xrn1-sensitive unstable transcripts (XUTs) (van Dijk et al., 2011). This polymerase, its transcripts and their processing are the main focus of this body of work.

These RNA polymerases are conserved in all eukaryotes and share many features with prokaryotic RNA polymerase (Cramer, 2002). All three eukaryotic polymerases share common components but also have unique protein

subunits; of note is the large subunit of RNAPII, Rpb1 (Rpo21p in yeast). This protein subunit has a carboxy-terminal domain (CTD) that is important for regulation of its activity and will be discussed later 1.2.

RNA polymerases proceed through three phases during transcription; initiation, elongation and termination (review Shandilya and Roberts, 2012).

1.1.1 Transcription Initiation

To initiate transcription gene specific factors bind to promoter sequences and distal enhancers on the DNA. These promote the recruitment of general transcription factors to form the pre-initiation complex, see Figure 1B, (Shandilya and Roberts, 2012), (Grünberg and Hahn, 2013), (Haberle and Stark, 2018). The polymerase begins transcription and the complex is now the "initially transcribing complex" (Figure 1C). Initially, abortive initiation only produces short RNA transcripts before dissociating from the template. Only after 25-27 ribonucleotides have been polymerised does RNAPII become stably associated with the DNA and initiation proceed to elongation; escaping the promoter.

The rate of transcription initiation is one of the determinants of transcript abundance, not all initiations lead to full length, sense RNA transcripts.

1.1.2 Transcription Elongation and Pausing

The RNAPII breaks the shackles of the initiation factors and proceeds down the gene elongating the nascent RNA strand as it goes (Imashimizu et al., 2014) (Figure 1E). Even having accomplished promoter escape the polymerase often pauses transcription at the 5' end of genes (Churchman and Weissman, 2011). Meanwhile the 7-methylguanisine cap is placed on the RNA, and factors bind to the cap (Figure 1D).

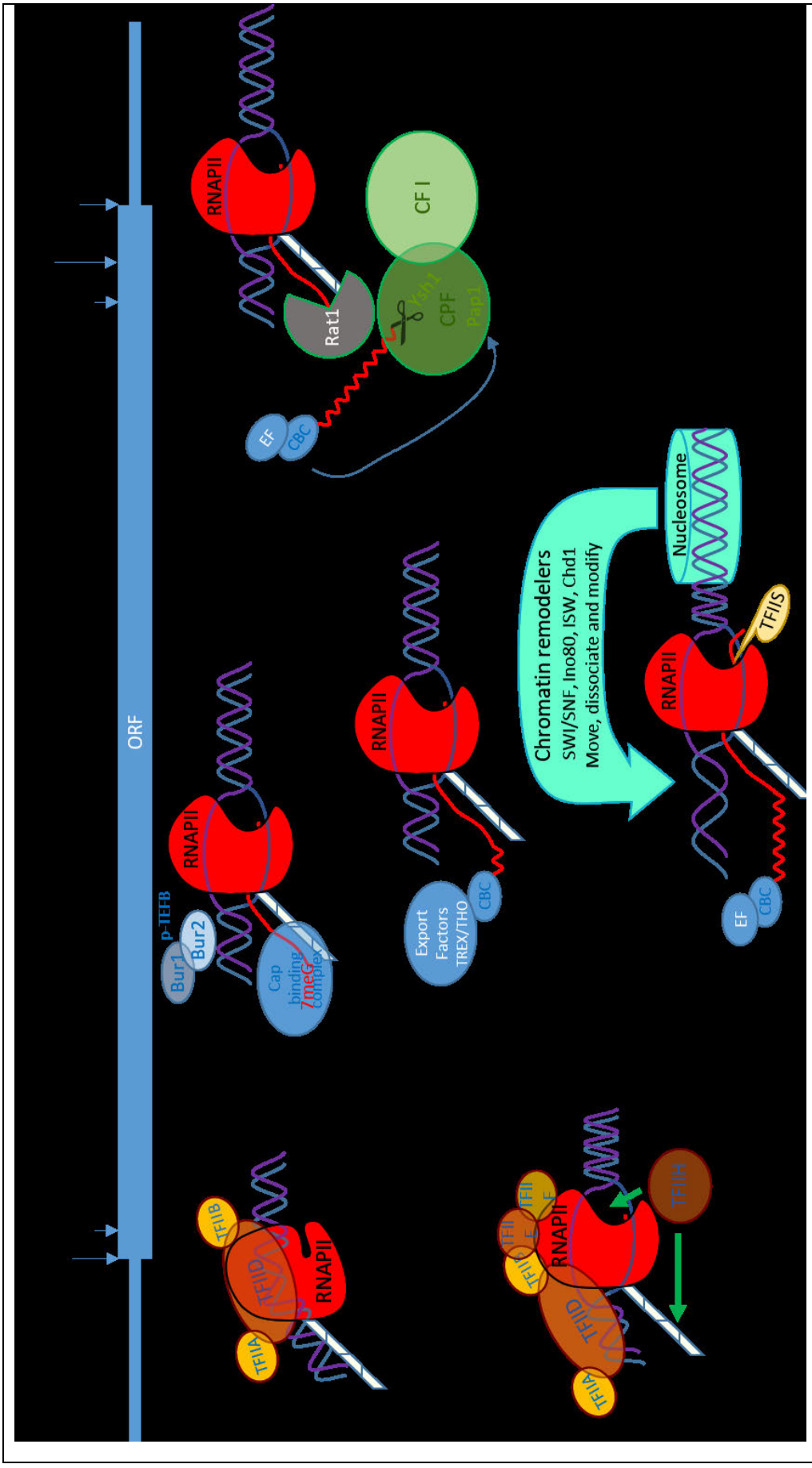


Figure 1 Transcription Cycle

Figure 1 Transcription Cycle
Some of the factors and processes involved in gene expression; those important for this thesis are discussed.
A) DNA template. **TSS**; is the Transcription Start Site, **AUG**; the first codon in the RNA, **stop**; the translational stop codon, **Poly-A-site**; the polyadenylation signal and finally the **Cleavage site**; the point at which the transcribing RNA is cleaved. The steps B to G are shown below at the approximate locations on the DNA.
B) Pre-initiation Complex. Factor TFIID binds the TSS in the promoter region, this induces a bend in the DNA. RNAPII, with the help of other transcription factors, docks onto the DNA.
C) Initially Transcribing Complex. TFIIH opens up the DNA allowing RNAPII to invade the DNA and bind a single strand. The polymerase undergoes several rounds of abortive transcription. DNA in front of the polymerase is hyper-wound ("scrunched"), a barrier to elongation.
D) Initially Elongating Complex. The polymerase has escaped the promoter and started productive transcription. The p-TEFB complex (Bur1p and Bur2p) phosphorylates the CTD of Spt5p; a region analogous to the CTD of RNAPII (Liu et al., 2009). The nascent RNA is 7-methyl guanosine capped and cap-binding complex (CBC) binds to it. This protects the RNA and also stimulates splicing (Lewis et al., 1996).
E) Elongating Complex. The polymerase transcribes the gene
F) Stalled/Paused/Backtracked Polymerase. The polymerase has encountered a block to transcription, in the example shown here a nucleosome, but can also be a pause for RNA processing or happen spontaneously. The DNA becomes over-wound and the polymerase stalls, it may then backtrack and if it does so the 3' end of the RNA protrudes out of the active site where the ribonucleotide normally binds. The tension is eased by nucleosome removal by chromatin remodelers such as Ino80p and chaperones (FACT) (Clapier et al., 2017), (Safina et al., 2017), and topoisomerase1 (Husain et al., 2016). TFIIS cleaves the RNA at the active site creating a new 3' end in the correct position for elongation, which can now resume.
G) Termination of elongation. Transcription proceeds past the Poly-A-site, although elongation slows. This is recognised by the multi-protein termination factors Cleavage Factor I (CFI) and Cleavage and Polyadenylation Factor (CPF). Ysh1p, a component of CPF, cleaves the RNA, Poly A polymerase adds the poly(A) tail. Rat1p digests the 5' end formed by this cleavage, removing the RNA still attached to the polymerase. Once the exonuclease reaches the polymerase they dissociate from the RNA; this is the "torpedo" model of transcription termination. The polymerase can then start another round of transcription. This is aided if the gene is looped so its 3' and 5' ends are close.

To transcribe a gene the polymerase must unwind the DNA and displace nucleosomes in its path (Figure 1F). Nucleosomes must then be repositioned behind the polymerase to prevent RNA invading the DNA (R-loops), causing genome instability (Aguilera and García-Muse, 2012). Elongation factors associate with the polymerase and are vital to promote transcription by removing impediments, preventing dissociation and release stalled polymerases. Many of these factors bind to the CTD.

On encountering a block to transcription, polymerase will stop and backtrack, in the order of 5 to 18 nucleotides. The nascent RNA protrudes from the site where the ribonucleotide triphosphate should bind. TFIIS (Dst1p in yeast) is required to cleave the protruding transcript and stimulate the restart of transcription (Ubukata et al., 2003) (Figure 1F). Depletion of this protein provided evidence that nucleosomes are a major block to transcription (Churchman and Weissman, 2011). Backtracking often leads to, or is the mechanism by which, the polymerase pauses on the gene.

The polymerase proceeds by Brownian motion, with a mechanism analogous to a ratchet, reducing retrograde steps and so promoting forward translocation (Bar-Nahum et al., 2005). Anything that prevents backtracking aids its forward motion. Positioning nucleosomes on the DNA behind it is one such impetus and secondary structure forming on the growing RNA is another. Binding of proteins, such as the spliceosome, to the nascent RNA chain could also prevent backtracking if they bind close enough to the polymerase.

1.1.3 Transcription Termination

Most mRNAs have the conserved polyadenylation signal sequence (poly(A) signal) AAUAAA near their 3' end. RNAPII elongates past this region, but slows. The RNA is cleaved at the poly-A-site (Figure 1G) (note the distinction from the poly(A) signal), and the 3' end polyadenylated. The RNA downstream of the cleavage site is degraded by the 5'-3' exonuclease (Luo and Bentley, 2004). Before cleavage any process happening to the mRNA transcript is by definition co-transcriptional. The precise site of cleavage is not fixed, but will

conform to a few favoured sites downstream of the AAUAAA sequence (Rosonina et al., 2006).

1.2 The CTD Code

The RNAPII largest subunit, Rpo1, has an extended C terminal. This is essential for viability and must be situated close to the polymerase's RNA exit port (Suh et al., 2013), but the tail is presumed to be flexible and can extend to 65 µm. The main portion of the tail is composed of 26 almost identical copies of the heptad repeat - Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y_1S_2P_3T_4S_5P_6S_7$). Most deviation from this consensus exists in the first and last two repeats.

All 7 of these amino acids can be post-translationally modified; all but the prolines can be phosphorylated, T_4, S_5 & S_7 glycosylated and the 2 prolines isomerised (Zaborowska et al., 2016), although not all modifications are found in budding yeast, see Table 1. The enzymes that add or remove these modifications are listed where known.

Table 1: CTD repeat phosphorylations in *S. cerevisiae*

Amino acid & position	Modifying Enzyme	Modification Removing Enzyme	Function and factors associated with the modification
Tyr1	Slt2p[1]	Rtr1p Glc7p	Possibly prevents premature termination
Ser2	Ctk1p	Fcp1p	Elongation and Transcription termination
Thr4	unknown	Fcp1p	Chromatin remodelling by Ino80p Termination Post-transcriptional splicing
Ser5	Kin28p[2] Cdk8p	Rtr1p Ssu72p	Initiation, capping, elongation, splicing. Involved in pausing, either helping to establish a pause or to promote release from a pause
Ser7	Kin28p[2] Bur1p	Ssu72p	Factor binding profile similar to Ser5, but fewer splicing factors

All sources (Zaborowska et al., 2016) and (Yurko and Manley, 2018) (Harlen et al., 2016) except [1](Yurko et al., 2017) [2] Intriguingly the *KIN28* gene contains an intron. Proline isomerisation seems to modulate the effect of the phosphorylated Serines before them (Yogesha et al., 2014) (Gibbs et al., 2017).

CTD modifications are dynamic and change as the polymerase goes through the stages of transcription. This can be seen as localisation to different regions

of the gene by Chromosomal Immuno-Precipitation (ChIP) (Alexander et al., 2010a) and in exquisite detail by crosslinking studies (Milligan et al., 2016). Different repeats can have different modifications or none at all (Suh et al., 2016), (Schüller et al., 2016). Furthermore, these modifications cluster to neighbouring repeats and different phosphorylations can even be found within the same repeat. The enzymes performing the modifications are recruited and activated by the transcription environment (Aitken et al., 2013)

The CTD binds proteins to the RNAPII at the appropriate stage; a landing pad, or possibly a fishing line. Interacting factors pulled down by immuno-precipitation (IP), using antibodies to specific phospho-modified forms of the CTD have been compiled (Harlen et al., 2016).

Of most relevance to the publications described in this study are the two abundant modifications, phosphorylation of Ser2 and Ser5. Ser5 is phosphorylated towards the transcription start site and, intriguingly, near the 3' ends of introns in budding yeast (Alexander et al., 2010a) and 5' end of introns in mammalian cells (Nojima et al., 2015). From IP studies it is known that phosphorylation of this serine promotes binding of transcription factors, mRNA capping factors and elongation factors, but also some splicing factors, particularly components of the U1 small nuclear RiboNucleoProtein (snRNP) (Harlen et al., 2016). This promotes co-transcriptional splicing.

Phosphorylated Ser2 by contrast interacts with factors involved with transcription termination and RNA export from the nucleus

Thr4P (review (Yurko and Manley, 2018) associates with the termination factor Rtt103p but, unusually among CTD phospho-residues it does not bind splicing factors. Its phosphorylation profile is also intron dependant; in intron-containing genes it has a bimodal distribution (Figure 2), whereas in non-intronic genes it only peaks after the poly-A-site. This could be a signal for either post-transcriptional splicing or failed splicing. Replacing a portion of the CTD's threonines with valine results in a splicing defect in 38% of transcripts.

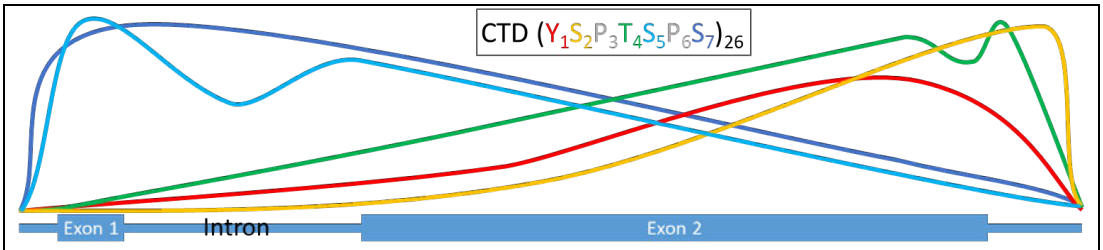


Figure 2: The CTD Code
Schematic graph showing how the phosphorylation state of amino acid residues in the CTD may vary along the length of an intron-containing gene. Of particular note is phosphorylation of Ser2 (yellow) rising towards the poly-A-site and of Ser5 (blue) indicating the promoter, transcription start site and rising over the 3'SS, these are both points where RNAPII may pause. References as Table 1.

1.3 Splicing

All protein coding mRNAs consist of 3 parts: the 5' untranslated region (UTR), the coding region and the 3'UTR. Most eukaryotic coding regions contain portions within it that do not code for a protein, these are introns, and need to be removed from the primary transcript before translation. Usually the removal will produce a continuous coding transcript by assembling the coding regions, exons, in-frame to form the mature mRNA. This process is akin to film editing where the frames required are spliced together to form a narrative. By analogy, this RNA editing is also called splicing. RNA splicing is found in all eukaryotes but not in prokaryotes, with the exception of the parasitic Group I and Group II introns.

The publications in this work focus on nuclear pre-mRNA splicing as performed by the large multicomponent ribo-nuclear protein complex, the spliceosome. The intron is identified by 3 sequences, the splice sites (SS); one at the 5' end of the intron (5'SS), one at the 3' end of the intron (3'SS) and one within the intron called the branch site (BS).

1.3.1 Spliceosomal Splicing in Yeast

The yeast *S. cerevisiae* can splice pre-mRNAs but, compared to other eukaryotes, the set of transcripts needing splicing is very much reduced. Only about 250-306 transcripts contain introns (Schreiber et al., 2015) out of five and a half thousand protein coding genes in total. Removal of only 3 of the 90

introns tested had a significant effect on growth in optimum conditions (Parenteau et al., 2008). Adjusting the level of expression and careful choice from among potential splice sites used to design artificially intronless genes restored growth. However, this study found that more of the introns were required under stress conditions.

When this study was extended to ribosomal protein genes (RPGs), it was found that maintaining expression levels was again an important function of the introns (Parenteau et al., 2011). This was particularly true between pairs of genes coding for the same protein. These paralogous genes are almost identical in coding region but differ in intron sequence but not location. Removing one of the pairs of introns changes both its and its paralogue's expression. As before, growth in rich media was unaffected, but removal of introns was deleterious for growth in the presence of drugs targeting protein production. The complete intron deletion experiments show introns are necessary to control protein levels, particularly to decrease ribosomal protein production in starvation conditions (Parenteau et al., 2019).

Evidence from global mRNA sequencing studies in yeast also reveals that there is little alternate splice site usage; mainly alternative 3'SSs and some exon skipping in the few double intron containing genes (Schreiber et al., 2015). Alternative splicing appears to be more a way of controlling expression rather than alternative protein production as most alternative site usage results in non-functional proteins. There are some instances of alternate protein production; *PTC7* intron retention adds a transmembrane domain (Juneau et al., 2009), *SRC1* alternate 5'SS usage produces a truncated, but still functional protein (Davis et al., 2000).

Although there is not the complexity or ubiquity of splicing in *S. cerevisiae*, compared to other eukaryotes, intron removal is still an important process for the organism. Although there are only a few intron-containing ORFs, these account for 27% of RNAPII transcription due to their high expression rate (Ares et al., 1999). It is possible that the introns boost the expression of their host genes (Le Hir et al., 2003), (Rose, 2019) but control of expression would be a

better description. A few introns contain snoRNAs, which maybe the intron's purpose (Hooks et al., 2016). Also some are known to regulate their transcript's abundance via specific factor binding (Vilardell and Warner, 1997).

1.3.2 Spliceosomal Introns in Budding Yeast

Introns in *S. cerevisiae* are also simpler than in higher eukaryotes and even other yeasts. The splice sites conform better to consensus sequences and are longer than those found in higher eukaryotes (Figure 3). The length of the intron is, conversely, much less than those common in metazoans, the longest just over 1000 nucleotides (nt) and almost all less than 600 nt. The relation between intron length and expression adds nuance to the observation, noted earlier, that intron-containing genes are highly expressed. Only genes with intron lengths, 300 to 600 nt long, are highly expressed. These are primarily RPGs or encode ribosome associated proteins.

Unlike many other eukaryotes, *S. cerevisiae's* introns are identified by the splicing machinery and spliced out. Other eukaryotes, with much longer and more abundant introns, splicing is believed to be identified via exon definition; splicing factor binding to a 5'SS stimulates the recognition of the upstream intron's 3'SS, mediated by a bridging complex spanning the shorter exons of higher eukaryotes. Initial and terminal exons require associations with the capping and termination complexes respectively to efficiently splice. Thus in non *S. cerevisiae* organisms the exons are identified and retained (Conti et al., 2013) leading to a less linear intron removal sequence (Drexler et al., 2020) although the order of splicing out intron seems to be consistent in transcripts and all introns in a transcript tend to be all efficiently or inefficiently spiced (Reimer et al., 2021).

The importance of RNA splicing to yeast, its comparative simplicity, the early availability of a complete genome sequence, the large array of mutant and deletion strains and a spliceosome that is conserved across eukaryotes makes *S. cerevisiae* an ideal model organism for studying the basics of splicing.

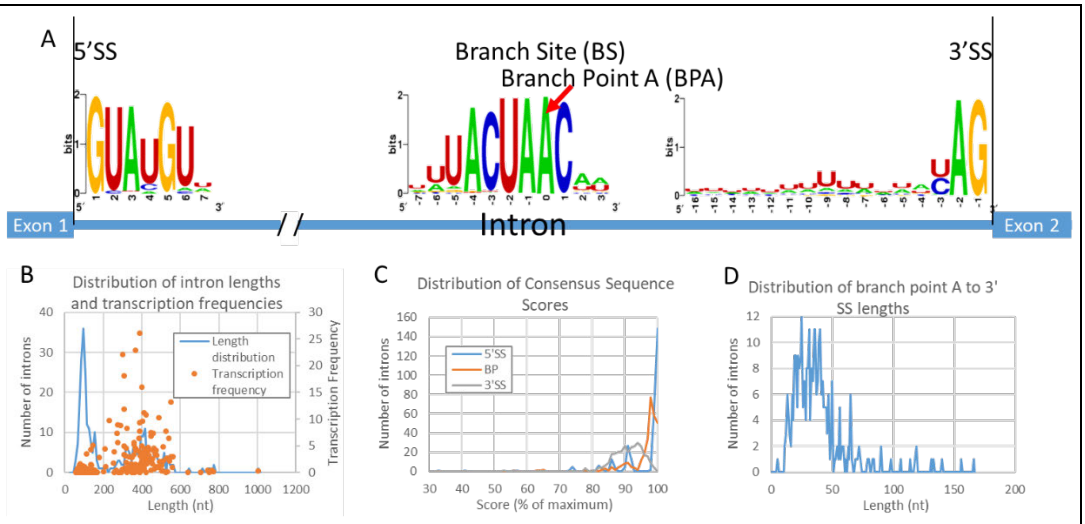


Figure 3: Anatomy of an *S. cerevisiae* intron

A) Sequence logos of the consensus sequences of the intron. The height of each base is proportional to the confidence of there being a consensus at that position. For positions where there can be more than one base they are shown, with the size representing the proportion of each base found at that position (Schneider and Stephens, 1990). The G at the beginning of the intron is conserved as are the final A & G of the intron. The branch site (BS) UAAC is very well conserved, with the final A in UA**A**C, the branch point adenosine (BPA), being absolutely conserved.
B) Length of introns. The distribution of intron lengths, apart from a few outliers, is bi-modal, with peaks of around 100 nt and another broader peak centred at about 400 nt. Red dots indicate the estimated transcription frequencies plotted on the second y axis against intron length for most of the intron-containing genes (García-Martínez et al., 2004). Transcripts with short introns are poorly expressed, while those in the larger peak are generally well expressed, ribosomal protein genes are in this peak. The outliers are poorly expressed.
C) Distribution of the strengths of the consensus sequences. The strengths are calculated using the algorithm used to create the sequence logos. The scores for the 5'SS are generally very high as only a few bases vary from the consensus; these are the lower peaks. The scores for the branch site are also high. The 3'SS has much lower scores, as only the AG is tightly conserved, the preceding polypyrimidine tract is not conserved.**D)** Length of the intron between the branch point and the 3' end of the intron. The vast majority of introns only have a short region of 10 to 50 nt between the BPA and the end of the 3'SS.

1.4 Spliceosome Assembly and Mechanism

Splicing by the spliceosome proceeds by 2 transesterification reactions; catalytic steps 1 and 2 (Figure 4). These steps are tightly controlled, with the 5'SS and BS consensus sequences recognised at least twice, redundant recognition of these sites helps prevent miss-splicing. The 5'SS is identified

first by the snRNPs U1 and then by U6. The BS is recognised by Msl5p (also known as Branch Point Binding Protein), and then by base-pairing with U2 snRNA. The 3'SS & polypyrimidine tract are less well monitored and the 3'SS is determined mainly by its proximity to the branch point; usually any AG within 45 nt of the BPA is chosen, with an optimum of about 25 nt (Figure 3D). Interestingly this is a physical distance, not a length of nucleotides. Secondary structure that brings a distant 3'SS closer to the BS promotes use of that 3'SS (Meyer et al., 2011).

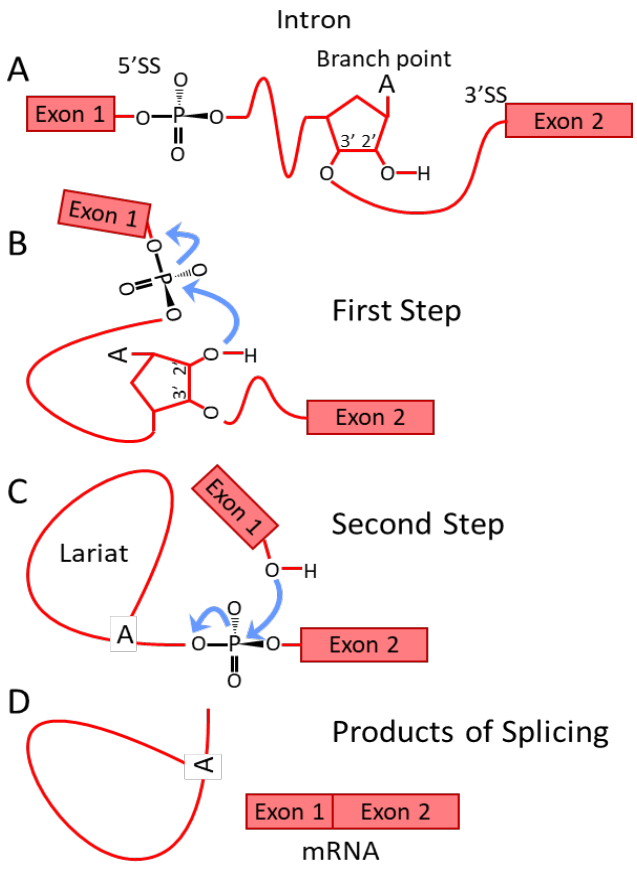


Figure 4: Splicing Schematic
A) The primary transcript before it is assembled into the spliceosome
B) The intron is looped to bring the 5' end of the intron close to the BPA of the intron. In the first step of splicing the 2' ribose oxygen nucleophilically attacks the phosphate between the first exon and the intron, breaking that phosphodiester bond and forming a 5'-2' phosphodiester linkage to the ribose of the branch point adenosine. This loops the larger portion of the intron into the lariat structure.
C) In the second step the OH at the 3' end of the first exon nucleophilically attacks the bond between the intron and the downstream exon resulting in the exons being joined and the intron excised.

The spliceosome is assembled in a stepwise manner (Figure 5) ensuring that each step is completed before the next can take place. The snRNAs play a vital role in the spliceosome, so much so that the spliceosome is largely a ribozyme co-ordinating two Mg^{2+} ions at the active site (Figure 6). The snRNAs bring the intron sequences together in the right way and at the right time for both steps of splicing to occur. Generally, the role of proteins is to stabilise the RNA structures and interactions and, in metazoans, direct alternative splicing.

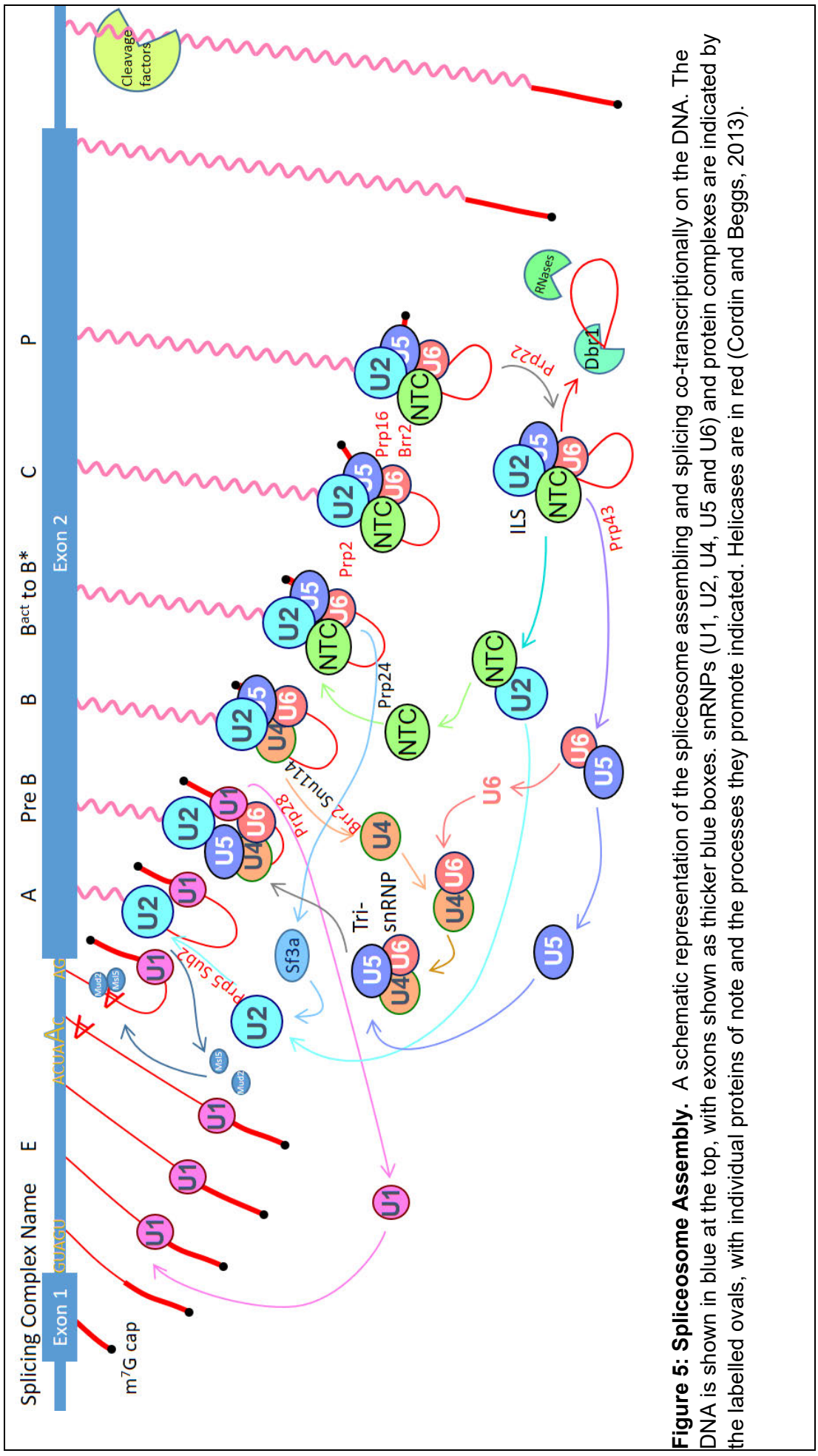


Figure 5: Spliceosome Assembly. A schematic representation of the spliceosome assembling and splicing co-transcriptionally on the DNA. The DNA is shown in blue at the top, with exons shown as thicker blue boxes. snRNPs (U1, U2, U4, U5 and U6) and protein complexes are indicated by the labelled ovals, with individual proteins of note and the processes they promote indicated. Helicases are in red (Cordin and Beggs, 2013).

Prp8p (component of U5 snRNP) is a particularly important spliceosomal protein. This protein has been called the "heart of the spliceosome" (Grainger and Beggs, 2005). It is conserved in sequence and structure to the reverse transcriptases of the parasitic Group II introns (Dlakić and Mushegian, 2011). This protein's ancient evolutionary origin points back to the first introns, with the protein encoded within the parasitic intron. Prp8p is crucial to splicing. It can be crosslinked to the 5'SS, BS and 3'SS and undergoes important conformational changes during splicing. A domain, the α finger, extends into the active site of the spliceosome at both catalytic steps (Fica and Nagai, 2017), (Shi, 2017), (Plaschka et al., 2019).

The spliceosome is highly dynamic and very complex. There are detailed and exhaustive catalogues of the splicing factors as determined by mass-spectrometry analysis (Fabrizio et al., 2009), (Will and Luhrmann, 2011), (Chen and Cheng, 2012) and the spliceosome database (Cvitkovic and Jurica, 2013). The helicases that harness ATP to aid conformational changes in the spliceosome and proofread splicing are highlighted in Figure 5 (Chang et al., 2013), (Cordin and Beggs, 2013), (De et al., 2016). Many of these helicases interact transiently with the spliceosome in order to check and aid in a particular step. It has been proposed that they pull on the protruding RNAs to draw the pre-mRNA into the active site (Hang et al., 2015).

The recently published cryo-electron microscopy (Cryo-EM) structures of the spliceosome are invaluable (for reviews (Shi, 2017), (Plaschka et al., 2019)). The structures have provided important insights and proved much that was inferred by genetic and biochemical techniques.

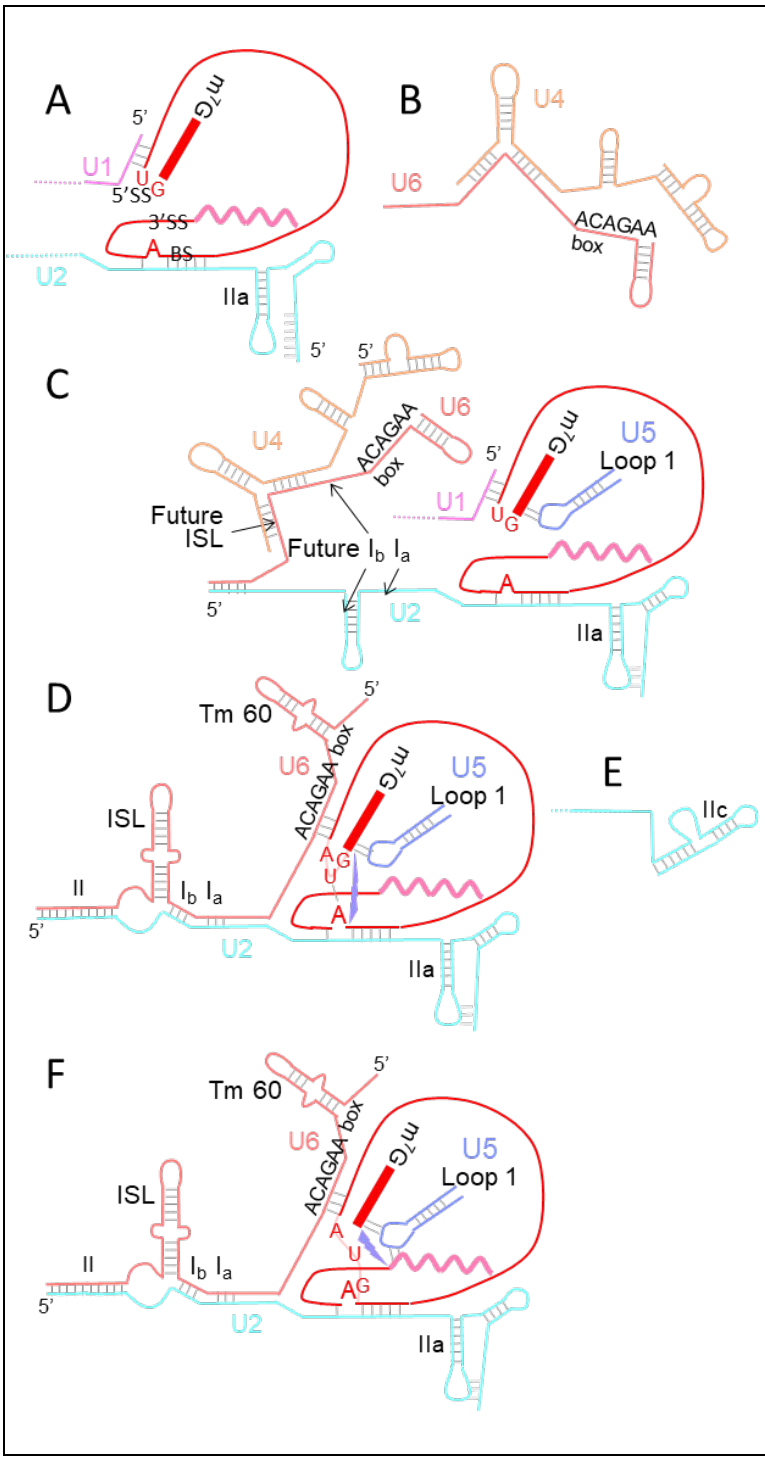


Figure 6: The Spliceosome is a Ribozyme
The messenger RNA is shown in red and the pink wavy line (exon 2). All other snRNAs are labelled in their colour. RNA features are named in black text. ISL = intra-molecular Stem Loop.
Helices Ia and Ib and Helices IIa and IIc
A) U1 and U2 bound and looping back the intron, the BPA is not base paired and bulges out
B) The U4 and U6 base-paired together in the Tri-snRNP.
C) RNA in Pre-B spliceosome. U5 contacts the exons, but U6 has not yet unwound from U4 and replaced U1.
D) B* spliceosome, U1 and U4 removed.
E) U2's stem and loop IIa may ordinarily form a pseudo knot with its own 3' end (shown in D). This may flip to an alternative IIc structure to initiate catalysis.
F) Complex C. The 2nd step is about to occur.

1.4.1.1 To the Commitment Complex (Complex E)

The 5'SS is recognised by base pairing to the very 5' end of U1 snRNA - me_3^{277}GpppAUACψψ, the underlined region being the reverse complement of the first 4 bases of the 5'SS (ψ is pseudouridine and me_3^{277}Gppp is the cap)

(Lerner et al., 1980). This interaction rather loosely defines the 5'SS region, at least in higher eukaryotes, allowing some flexibility in 5'SS choice later (Aebi et al., 1987). Msl5p (Bbp) recognises the BS. Mud2p (the yeast equivalent of U2AF65), previously thought to bind exclusively to the polypyrimidine tract, has been observed to interact along the entire length of the intron (Baejen et al., 2014). These interactions define the commitment complex (complex E), which, once formed, commits the pre-mRNA either to be spliced or degraded (Seraphin and Rosbash, 1989).

1.4.1.2 Assembling the A complex

U2 snRNP displaces Msl5p and Mud2p and occupies a position over the BS and polypyrimidine tract. U2 snRNA base pairs to the pre-mRNA's BS via the interacting sequence GUAGUA to form the branchpoint helix. The BPA is bulged out between the two bases GU underlined as it has no complementary base to pair with (Figure 6A). This is stabilised by the helicase, Prp5p, which leaves the spliceosome once the branchpoint helix is formed, thus allowing the next stage to proceed (Liang and Cheng, 2015). Prp5p and its interacting partner Hsh155p of SF3b complex, which is part of the U2 snRNP, are major determinants of correct BPA selection (Tang et al., 2016). This forms the A complex or pre-spliceosome. The SF3a complex, containing the proteins Prp9p, Prp11p and Prp21p, is also part of U2 snRNP at this stage.

1.4.1.3 B Complexes

The tri-snRNP, Figure 6B, consisting of the snRNPs U4, U5 and U6, joins, to form the relatively unstable pre-B complex (N.B. in older publications, the pre-B and B complexes could not be distinguished). U4 and U6 are still extensively base-paired together at this point (Figure 6C).

After the tri-snRNP binds; U1 is displaced from the 5'SS by the helicase Prp28p. Conversion of the B to B^{act} complex involves addition of the NineTeen Complex (NTC) and Brr2p, a helicase and U5 snRNP component, which disrupts base pairing between U4 and U6, allowing the pre-mRNA's 5'SS to hybridise with the ACAGA(GA) box motif of U6 (named after its sequence). The underlined bases are complimentary to the 5'SS region, leaving the first 3

bases of the intron free to engage in the splicing reaction. U6 also now forms extensive base pairing with U2, of which helix Ia and Ib are part of the spliceosome active site. As U2 is bound to the branch site and U6 to the 5'SS, these two regions of the pre-mRNA are brought into close proximity (Figure 6D).

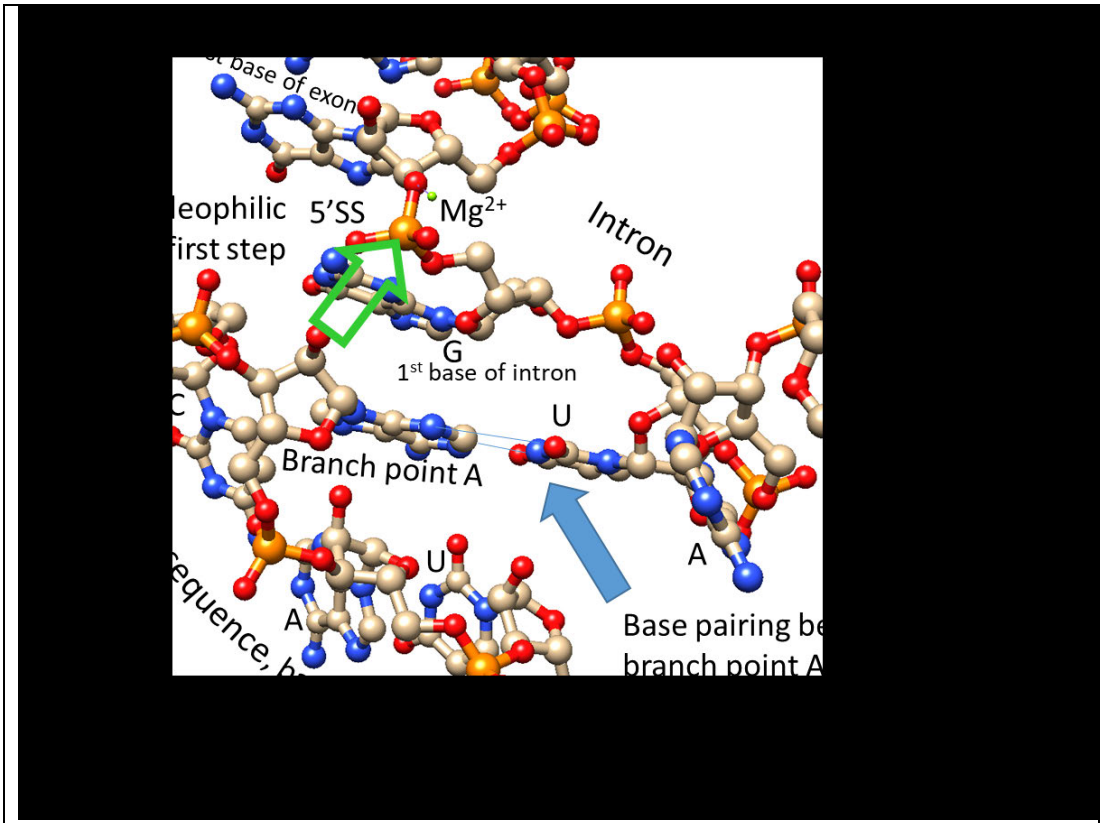


Figure 7: The Catalytic Centre Just Prior to the First Step of Splicing.
U2 and U6 conspire to bring the 2' oxygen of the BPA in close proximity to the 5' phosphate of the 5'SS initial G and a magnesium ion (Mg^{2+}) by the formation of the U2 stem IIc (Hilliker et al., 2007), (Perriman and Ares, 2007) (Figure 6E). The BPA is twisted out of the helix, as it does not bind to U2, making it available to base pair with the second base of the 5'SS, a U (Fica and Nagai, 2017). This helps to stabilise the close proximity of: the 2' oxygen of the BPA's ribose, the phosphate in between the exon and intron and a magnesium ion. This promotes nucleophilic attack from this oxygen to the phosphorus making a 2' to 5' phosphodiester bond, creating the lariat and breaking the 5'-3' bond between the upstream exon and the intron. The phosphate is less likely to circularise as the target 2' O is closer, and any other new phosphodiester bonds will not release the upstream exon. The first step of splicing is favoured if the exon springs away from the active site once released.

U5 loop1 interacts with the first exon close to the 5'SS (McGrail and O'Keefe, 2008) (Figure 6D). Components of the NTC, the NineTeen Related complex (NTR) and RES (REtension and Splicing) complex, displace the SF3a factors

from U2. The Prp2p helicase promotes further conformational change to form the B* complex, primed for the 1st catalytic step (Figure 7).

1.4.1.4 C and C* Spliceosome

On completion of the first catalytic step, some additional factors required for step 2 arrive to form the C complex. Of note is the helicase Prp16p, the instigator of rearrangements for the second step. The spliceosome is remodelled to bring the 3'SS and the 3' end of the upstream exon into proximity so the second step can take place (Complex C*). Recent cryo-EM structures reveal that the BPA of the lariat is involved in recognising the 3'SS via reverse Hoogsteen pairing (Figure 8). The two exons are held close to each other by U5 loop 1, now interacting with both exons (Figure 6F). The second step factor, Prp18p, may help bring 3'SSs further away from the BPA than optimal closer to the active site. The second catalytic step of splicing can now take place (Figure 8)

1.4.1.5 P Spliceosome, Intron/Lariat Spliceosome and Recycling

Structurally there is little to distinguish the post second step spliceosome (P complex) and the pre (C*), other than the exons being ligated together. The mRNA is released by Prp22p (another helicase), leaving the excised intron lariat with the spliceosome, the Intron/Lariat Spliceosome (ILS). The intron is released by Prp43p helicase, debranched by Dbr1p and degraded. Prp43p also functions to release stalled spliceosomes that cannot complete splicing at earlier stages (see chapter 7). The spliceosomal components are recycled and the tri-snRNP reforms.

1.5 Transcription and Splicing

RNA processing events that occur during transcription are referred to as being co-transcriptional. The 5' cap added to the RNA is uncontroversially co-transcriptional (Cho et al., 1997) (McCracken et al., 1997). Factors that cleave the transcript and terminate transcription also act co-transcriptionally. Of more debate is mRNA splicing, although strong evidence is provided by the electron

micrograph on the front of this thesis (Beyer and Osheim, 1988), (Proudfoot, 2000) showing co-transcriptional splicing in Drosophila.

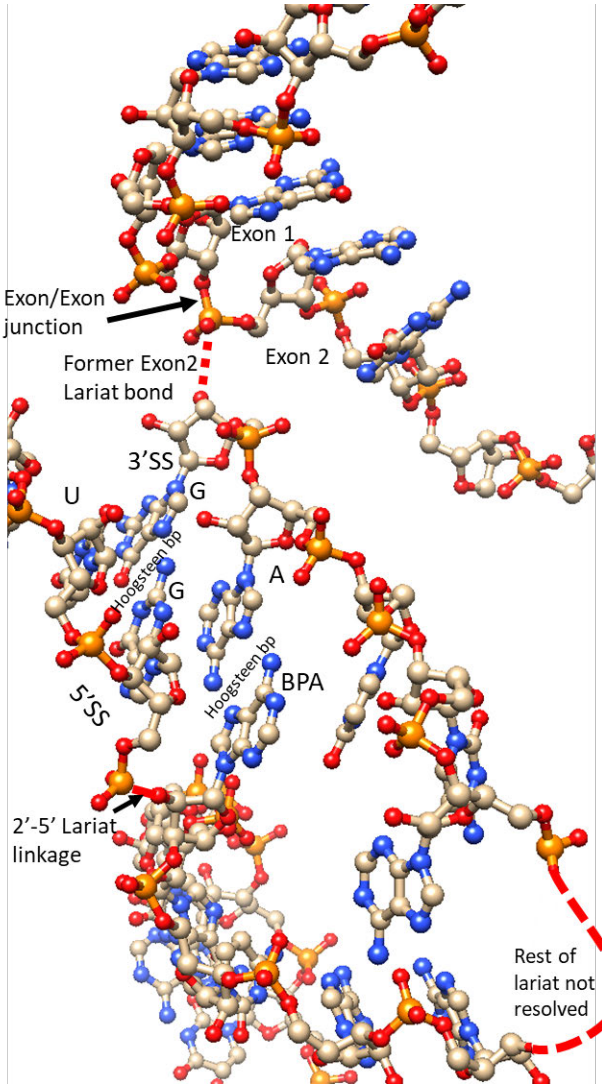


Figure 8: Spliceosome immediately after second step of splicing
The lariat is to the left and lower portion of the figure, the mRNA upper right.
None of the existing cryo-EM structures of the C* complex show the 3'SS or exon 2 (Fica et al., 2017), (Zhang et al., 2017), (Bertram et al., 2017) or (Yan et al., 2017). Presumably these RNA regions do not resolve as they are being pulled by Prp22p, scanning for the AG dinucleotide at the 3'SS. Some insight can be gathered from examining the P complex (Wilkinson et al., 2017)
The BPA forms a reverse Hoogsteen base pair with the penultimate A of the intron and the 5'SS G forms another with the last base of the intron G. These pairs are stacked. This is probably how the 3'SS is recognised. Possibly the formation of the 2 base pairs is enough to signal that the 3'SS has been found and draw the phosphodiester bond close enough to the 3' OH of exon 1 and the Mg^{2+} ion for the nucleophilic attack from the OH to the phosphorus to take place, joining the exons together and releasing the lariat. The region between the BPA and 3'SS is unresolved in the structure and therefore unconstrained by the spliceosome. This will allow secondary structure to form, reducing the effective distance between these 2 sites (Plass et al., 2012)

Co-transcriptional spliceosome recruitment is advantageous in that; protein binding can protect RNA from degradation, spliceosome components can accumulate near intron-containing transcripts making splicing faster, splicing and transcription can communicate and modulate each other. The splice sites

emerge one at a time from the polymerase so components can be loaded onto the pre-mRNA in the correct order, simplifying the task of spliceosome assembly and aiding fidelity.

1.5.1 Splicing Organelles

The assembly and disassembly of the spliceosome involves cycling of components as they are added, removed and recycled (Figure 5). Concentrating the components near the locus of expression should improve the efficiency of spliceosome assembly. Diffusion rates measured in human cells for several snRNPs are lower than would be expected if they diffuse freely (Huranová et al., 2010). Recent work on liquid–liquid phase separation (LLPS) (Courchaine et al., 2016) provides a mechanism for how membrane-less organelles can be created and maintained. Examples are: nucleosomes, Cajal bodies and possibly nuclear speckles in higher eukaryotes, which is relevant as they often contain transcription and splicing complexes (Spector and Lamond, 2011).

1.5.2 Evidence of Co-transcriptional Splicing

The two processes of transcription and splicing are often considered to be separate events, particularly as they can be separated *in vitro,* and earlier research suggested that co-transcriptional splicing was rare (Tardiff et al., 2006). ChIP experiments detected splicing factors cross-linked to DNA (Kotovic et al., 2003), (Görnemann et al., 2005), (Tardiff et al., 2006) at biologically relevant locations, indicating that the DNA and RNA must be very close when the factors bind.

Studies using synthetic reporter genes (Ribo1, (Alexander et al., 2010b) see chapter 2), where transcription can be induced very rapidly, show splicing occurring within minutes of transcription starting. *In situ* hybridisation shows there is an absence of intron signal away from the presumed site of transcription (chapter 2). Modelling of transcription and splicing of this reporter suggested that co-transcriptional splicing of both steps is the most efficient means of producing mRNA and best matches the data (Aitken et al., 2011).

Furthermore, the half-life for step 2 splicing is in the order of 5.5 seconds (once transcription has been induced for over 11 minutes), but 110 seconds in the first round of splicing. This indicates two things: that it takes time for the splicing factors to accumulate and secondly the value of co-transcriptional splicing once they are in the locale.

Genome-wide techniques such as "nascent RNA-seq" (Harlen et al., 2016) and the related "NET-seq" (Native Elongating Transcript - sequencing), (Churchman and Weissman, 2011), (Nojima et al., 2015), offer a more direct indication of the location of polymerases on genes. In these studies, RNA associated with chromatin is prepared and an oligo ligated onto the 3' end. Sequencing from this oligo provides not only information on the 3' end of the transcript (taken to be where the RNAPII was at the time of sampling), but also whether the RNA is spliced or unspliced. For example, a spliced RNA with a 3' end within 100 nt of the 3'SS would be taken to indicate that splicing occurred co-transcriptionally. However, as the RNA is only associated with the chromatin rather than the polymerase this is an assumption that may not be justified as polyadenylated RNA is associated with this fraction (Drexler et al., 2020).

Single Molecule Intron Tracking (SMIT), (Carrillo Oesterreich et al., 2016) is a related technique in that RNA is prepared from chromatin and harshly washed to capture RNAs associated with RNAPII rather than by IP. A telling observation is that very few 3’ ends mapped to within introns in the SMIT data for RPGs, suggesting that the polymerase transits the intron fast compared with the downstream exon.

These techniques, along with metabolic labelling of nascent RNAs using 4-thiouracil (4tU) followed by purification and sequencing (Barrass et al., 2015) (chapters 3 and 4), were combined to produce overall estimates of the degree of co- versus post-transcriptional splicing (Wallace and Beggs, 2017). This study will be discussed in more detail in chapter 4. Briefly, there are two groups of intron-containing transcripts; those that are efficiently transcribed and spliced within 100 nt downstream of the 3'SS, and those that are produced

slowly and spliced further from the 3' end of the intron, a few even post-transcriptionally spliced. The first group contains almost all the intron-containing RPGs.

Processes involved in 3'end formations termination in *S. cerevisiae* could remove signals leading to co-transcriptional splicing. If Nuclear polyA binding protein (Nab2p). is depleted there is a defect in transcription termination (see chapter 4). This leads to run on transcription into downstream genes. The introns in these genes are normally co-transcriptionally spliced, however, the run on transcripts are not (Alpert et al., 2020).

Lessons learnt in yeast seem also to apply in mammals, the introns may be only spliced several kilobases from the 3'SS, (Drexler et al., 2020) but 75% are spiced within 300 nt of a 3'SS (Reimer et al., 2021). The interaction in mammalian cells between splicing and 3' end processing is present, but poorly spliced transcripts are associated with poorer 3' end cleavage.

1.5.3 Transcriptional Pause for Splicing

Given the prevalence of co-transcriptional splicing does the RNAPII pause to allow this to happen? Does it happen during transcription of all genes all the time? It is also not clear what form any pause might take. Does the polymerase stop at a point for splicing, wait until it receives signals that splicing has occurred and resume transcription? At the other end of the scale from this, is the polymerase just a little less progressive and stutter along the downstream exon at about the same time that splicing takes place? This is well reviewed in Neugebauer (2019) however, further work will have to be done, the answer, as usual, will most likely lie in-between the two models, but whatever the answer splicing must be able to influence polymerase's processivity.

Pausing, or slowing of polymerase, is consistent with the kinetic model of co-transcriptional splicing (Saldi et al., 2016). Although it is generally thought of as determining alternative splice site choice, changing the speed of transcription does alter splicing in yeast, and it appears that transcription

speed is tuned to optimise efficiency and fidelity of splicing (Aslanzadeh et al., 2018).

1.5.3.1 CTD Phosphorylation

ChIP studies with the inducible reporter Ribo1 (Alexander et al., 2010a) suggested that RNAPII builds up at the beginning of the second exon. This is indicative of the polymerase being paused there. ChIP using antibodies against CTD Ser5P also showed an increase at the same point; the paused polymerase is CTD ser5 phosphorylated. An elegant experiment, where splicing was restored to a BS site mutant by a complimentary U2 mutant, showed that this pause was splicing, not sequence, dependant.

The pause found by ChIP (Alexander et al., 2010a) occurred in a recently induced artificial reporter, so it is conceivable that the pause could help set up splicing on a newly induced gene. As the spliceosome will already be in place near a gene with well-established expression a pause may not be necessary. However, ChIP-seq shows this same accumulation very well on endogenous genes (Chathoth et al., 2014). Curiously, this paper also shows a distinct paucity of RNAPII over the BS, suggesting the RNAPII is particularly unimpeded at this site.

SMIT-like experiments, using tiling arrays rather than sequencing, also identified a transcriptional pause on average 250 nt downstream from introns (Carrillo Oesterreich et al., 2010). A change in RNAPII's behaviour was found in intron-containing genes with short exons, the polymerase slows elongation within 250 nt of the poly-A-site, apparently to ensure splicing is co-transcriptional. This study found that splicing is important for this pause corroborating the experiment where splicing is restored to a BS mutant mentioned previously (Alexander et al., 2010a).

In a later study Carrillo Oesterreich (2016) claimed that their SMIT data showed that splicing can happen extremely fast and that a pause was unnecessary for co-transcriptional splicing. However, their technique could not detect splicing intermediates, the species most likely to be associated with a

paused polymerase, so it would appear to the authors that RNA transitioned from unspliced to fully spliced almost instantly.

Nascent RNA-seq has also provided evidence of RNAPII build up in exons downstream of introns, interpreted as paused polymerases (Harlen et al., 2016). Moreover, most paused polymerases were Ser5 phosphorylated on their CTD.

1.5.3.2 Ubiquitination of RNA Polymerase II

The largest subunit of RNAPII can also be ubiquitinylated, this seems to be crucial for splicing-associated pausing (Milligan et al., 2017). Ubiquitin on residue K1246 of Rpo21p can restrict access to the DNA channel. If Bre5p, cofactor of the enzyme that de-ubiquitinates Rpo21p at this site, is deleted the polymerase pauses before the 3'SS (possibly over the BS), and spicing efficiency is reduced. If K1246 is mutated, pausing is reduced, even in the absence of Bre5p, and splicing becomes more post-transcriptional. Bre5p seems to be a major player in controlling elongation by influencing de-ubiquitination. It binds RNA produced by RNAPII with a preference for UUUG sequences in the downstream exon of intron-containing transcripts. Interestingly, deletion of *BRE5* or *UBP3,* its ubiquitin protease partner, is synthetic lethal with certain mutations in *PRP45,* an NTC component (Beggs lab unpublished results). A plausible model is that Prp19p, another NTC component and ubiquitin ligase, ubiquitinates RNAPII to promote pausing during splicing. After splicing completion, the Bre5/Ubp3 complex is recruited to de-ubiquitinate and release the paused RNAPII. The ubiquitin-like molecule Hub1p reduces splicing fidelity when bound (Chanarat and Svasti, 2020).

1.5.3.3 Chromatin and Nucleosomes

The distribution of nucleosomes alters between introns and exons, nucleosome positioning being much more defined in exons (Schwartz and Ast, 2010). The alternative histone H2A.Z localises over the introns of non-ribosomal protein genes (Neves et al., 2017). The nucleosome environment will modulate the elongation speed

The histones have different post translational modifications on their N-terminal regions, in particular H3K36me3 (Histone 3 Lysine 36 tri-methylated) is found in exons (Sorenson et al., 2016). A link between H3K36me3 has been found to splicing via Eaf3p, a protein that can interact with methylated H3K36. This protein Co-IPs with Prp45p and can possibly recruit the NTC to the spliceosome (Leung et al., 2019). H3K36me3 could place the NTC close to the second exon, ready to promote U4/U6 unwinding and start the splicing process.

1.5.3.4 Checkpoints

Prp5p provides additional insights (Chathoth et al., 2014), (I am a second author on this publication). The *prp5-1* mutation lies in motif I of the DEAD-box helicase Prp5p and confers temperature-sensitive growth. At permissive conditions splicing proceeds normally, but at non-permissive temperatures, spliceosome assembly stops with only U1 and U2 bound to the pre-mRNA. ChIP shows that RNAPII is stalled within introns, and that the CTD is Ser5 phosphorylated. This indicates that: pausing can occur within the intron, the pause is not evident under normal circumstances and also that Ser5P CTD is associated with a pause (Milligan et al., 2016). Cus2p is retained on U2 snRNP in the *prp5-1* mutant strain. If the *CUS2* gene is deleted this pause disappears and transcription is restored, but not splicing. It was proposed that Cus2p acts like a checkpoint factor monitoring Prp5p function (Chathoth et al., 2014). If Prp5p does not function normally Cus2p signals to the polymerase to pause, either to allow additional time for Prp5p to complete its task or to degrade the RNA. If it is not present to inform on Prp5p activity transcription proceeds oblivious to the state of splicing. Recently a genetic link and physical interaction from Prp5p to histone modification was identified (Shao et al., 2020), linking Prp5p to transcription.

A model based on all these studies is that a pause is difficult to detect because it is not localised, occurring in different locations and times in different cells in a population. A failed checkpoint can make a diffuse and stochastic pause

more defined and endure longer, thus making it more likely to be detected by current techniques.

1.6 RNA Degradation and Quality Control

The means of RNA degradation are many and varied (Parker, 2012), reflecting RNA's role as an intermediate; its transience is vital to maintaining metabolic flexibility. RNA levels are the product of the competing processes of transcription and degradation and this balance is an important factor in gene regulation (Timmers and Tora, 2018).

The RNA degradation processes important to this work are de-branching, and 5' to 3' and 3' to 5' degradation. Debranching is where the lariat product of splicing is un-looped by breaking the 2'-5' phosphodiester bond at the 5'SS to branch point junction. This is accomplished by Dbr1p (Chapman and Boeke, 1991), allowing the other methods of degradation to occur.

Degradation from the 5' end is performed by the exonucleases Rat1p and Xrn1p. De-capping must occur first and is the rate limiting step in degradation (Sohrabi-Jahromi et al., 2019). 3' end degradation is performed by the exosome, a large multi-component complex; removal of the poly A tail is a prerequisite for this to occur and often shortening of the poly(A) tail leads to de-capping and degradation.

Translation is quality controlled (QC) and defective transcripts eliminated; often these are the result of mis-splicing. Translation QC pathways in the cytoplasm are: no-go decay (NGD), nonsense-mediated decay (NMD) or its converse non-stop decay (NSD) (Shoemaker and Green, 2012). However, there is much degradation in the nucleus (Schmid and Jensen, 2018) and there is some evidence that splicing of intronless transcripts is a means to target transcripts for decay in the nucleus (Volanakis et al., 2013).

In the timing model of quality control (Libri, 2010) export and degradation are in competition (Tudek et al., 2018). Anything that delays removing RNA from the dangerous, RNase filled, environment of the nucleus increases the chance

of degradation (Peck et al., 2019). Slowing: escape from the promoter, release from an elongation pause, capping, splicing and 3' end formation all increase the likelihood of degradation. The timer is probably the time taken for an active degradation complex to be formed (Schmid and Jensen, 2018).

QC in splicing works on the same principals (Semlow and Staley, 2012). Prp16p is required to dissociate factors Yju2p and Cwc25p after first step using the energy from ATP hydrolysis. However, on a substrate that is slow to splice, Prp16p can hydrolyse ATP and remove the factors before step one has occurred (Koodathingal and Staley, 2013). This leads to a stalled spliceosome, which is released by Prp43p and the RNA degraded (chapter 7). Thus, the second step factor Prp16p proofreads the first step of splicing. The timer in this case is the time taken for Prp16p to bind, hydrolyse ATP and displace the factors. Rejection of the substrate is reversible until Prp43p dissociates the spliceosome (Koodathingal et al., 2010). This raises the interesting possibility that the spliceosome goes through multiple rounds of ATP hydrolysis, site rejection and reacquisition to find the optimal splice site (Semlow and Staley, 2012). Similar mechanisms have been demonstrated for Prp5p (Xu and Query, 2007) (Chathoth et al., 2014), (Liang and Cheng, 2015) and Prp28p (Yang et al., 2013) and could be a general feature of all the spliceosomal helicases.

The aim of RNA QC is to protect the cell from waste of resources as protein production is even more energy and resource intensive than transcription, as many proteins can be produced from one mRNA. Degrading the RNA is a cost worth paying, to borrow a phrase from the pharmaceutical industry "fail early, fail cheap".

The cost of this rigour is the potential to degrade many correct transcripts (Porrua and Libri, 2013), but this also adds another layer of control to gene expression. An instructive example is provided by depletion of the protein Dis3p (also known as Rrp44p), a core component of the exosome. On depletion the level of mRNA increases, indicating that a few valid RNAs are targeted for degradation and that QC is not a one-way street leading inexorably to decay. Some pre-mRNAs targeted for decay may be rescued by mutants

affecting degradation. The levels of these RNAs are increased and they may go on to be spliced (Gudipati et al., 2012). See also chapter 5 for splicing microarray data showing that crippling the exosome leads to more complete splicing of some transcripts.

1.7 Small Nucleolar RNAs (snoRNAs)

snoRNAs are a class of RNAs that direct modification of other RNAs (Piekna-Przybylska et al., 2007). The modifications are mainly methylation and pseudouridylation but can include target cleavage. The targets of this modification are the rRNAs but U2 snRNA is pseudouridylated by snR81 (Ma et al., 2005).

snoRNA transcription is either mono-cistronic by RNAPII/III, poly-cistronic by RNAPII or within an intron. In all cases the snoRNA is processed by cleaving out the RNA at a hairpin by the dsRNA endonuclease Rnt1p (RNase Three) (Hartman et al., 2013) and end-trimming by the exosome or Rat1p until the snoRNA's binding proteins are encountered. Intronic snoRNAs can be processed two ways, either released from the spliced and debranched intron (Ooi et al., 1998) or cleaved from the entire transcript (Giorgi et al., 2001). There are only seven introns with snoRNAs, three are second introns and in all but one the containing gene's protein product is ribosome associated.

1.8 Overview of Methods used

Many of the methods used in these publications are familiar in any molecular biology lab. Others are less familiar but have not been developed by this author. These methods will not be discussed here. Methods that have been developed, adapted and/or optimised by the author are expanded upon in this section. The exact procedural details of these methods are given in the included publications, particularly the two published in JoVE (Barrass et al., 2019), (Barrass and Beggs, 2019). This section gives an overview of how the techniques were developed over time with additional details to aid understanding.

1.8.1 Microarray

Microarrays were pioneered in the late 1990s for studies with *S. cerevisiae* (DeRisi et al., 1997), (Lashkari et al., 1997). DNA probes are printed in discrete spots onto glass slides. Several thousand such probes can be printed onto a slide (often termed a chip), enough for one probe specific to each ORF of *S. cerevisiae*. Later chips, using slightly different technology, could provide complete coverage of the yeast genome or every exon for *Homo sapiens*. This technique offered unparalleled insight into the transcriptome, ushering in the age of transcriptomics.

Fluorescently labelled cDNA produced by reverse transcription from total RNA extracted from the yeast is hybridised to the probes. Measuring the fluorescence bound to each probe spot provides an estimate of the amount of cDNA hybridised to each probe and therefore levels of mRNA in the sample.

Microarrays are not without their limitations. The linearity of the signal measurement is poor, and absolute levels of RNA abundance cannot be determined. Therefore, microarrays are comparative; mutant versus isogenic wild type or changes from a base in a time course. This is possible as RNA from paired samples prepared at the same time can be reverse transcribed with two different fluorescently labelled nucleotide analogues (Cy3 and Cy5) and hybridised on the same microarray chip. To avoid bias due to differences in dye efficiency and incorporation, the control and test dyes are swapped in different replicates. This comparative approach also does much to remove systematic biases, such as probe printing efficiency, reverse transcription efficiency, operator and procedural differences.

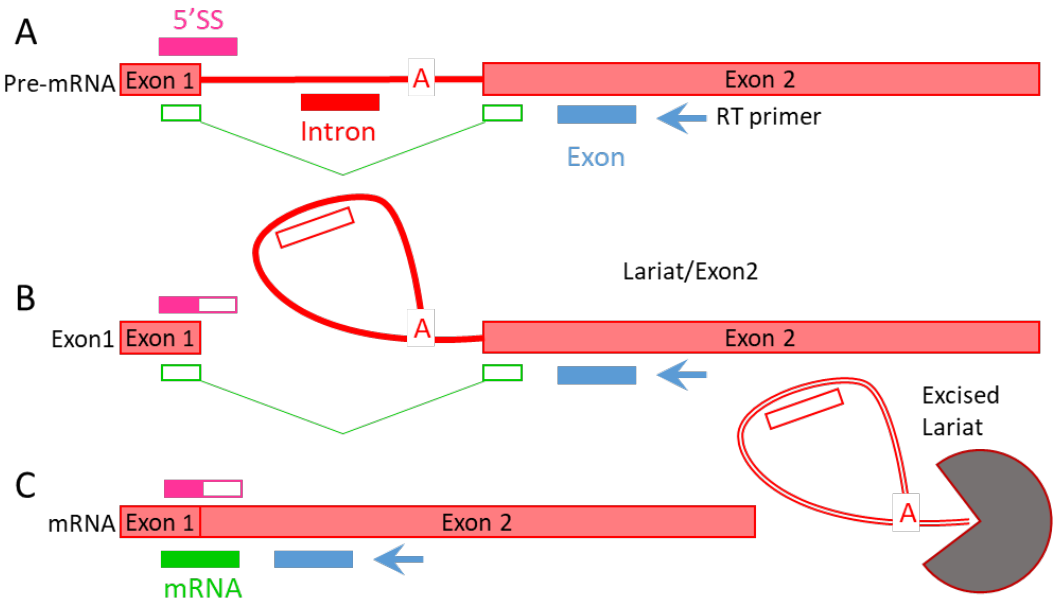
It was not long before microarrays were applied to assay splicing (Clark et al., 2002) (Barrass and Beggs, 2003). The work by the two labs overlapped but follows the same principles and with the same limitations. Details are provided in the publications bound in this thesis (chapter 5), but, briefly, probes were designed to bind regions diagnostic for splicing (Figure 9). Other probes were

to the intronic snoRNAs, a few other snoRNAs, the spliceosomal snRNAs (U1, U2, U4, U5 and U6) and a few intronless genes.

Probes were made for all introns known at the time of design, others were added as more were discovered. Re-designs were also necessary as not all introns were annotated correctly in the Saccharomyces genome database (SGD) (Cherry et al., 2012). Introns annotated up to and including release 62-1-1 (2009-02-18) of the SGD were included in updated microarrays. Of great help were the intron databases (Spingola et al., 1999), (Lopez and Séraphin, 2000). Where there were differences between sources, probes were also made to assay all permutations of exon/exon junctions that could result if all annotations were correct. The probes were printed onto glass slides in triplicate. An example of a microarray scan is shown at the beginning of chapter 5.

A thermostable reverse transcriptase was used to create the cDNAs to improve specificity, with the elongation reaction performed at 50-55^{o}C, improving enzymes allowed increasingly higher temperatures to be used.

Normalisation for splicing microarrays was performed by dividing the fluorescent intensity in each channel (Cy3 or Cy5), for each probe by the intensity in that channel for that gene's exon probe, see Equation 1. This was an attempt to normalise away gene expression differences. Without this normalisation, a transcript present in a higher amount in one sample versus the control would appear poorly spliced, merely because it is more abundant. This exon normalisation was only partially successful as exons will always be the most abundant sequences and the non-linearities inherent in microarray signals reduced the power of this strategy. More standard methods of normalisation, using global measures, were not suitable because the assumption of no overall change between test and control is not valid in a splicing microarray where intron retention will massively change the overall levels between the test and control.



D

	Pre-mRNA	Lariat /Exon2	Exon1	Excised Lariat	mRNA
5'SS	✓	×	×	×	×
Intron	✓	×	×	×	×
mRNA	×	×	×	×	✓
Exon	✓	✓	×	×	✓

Figure 9: Splicing microarray probes
A) Pre-mRNA; Probes against the 5'SS, intron and exon 2 will detect this RNA. The probe in the last exon was ideally within 100 bases of the intron with the RT primer as close as possible downstream (represented by an arrow in the second exon).
B) RNA after the first step of splicing; only exon 2 can be detected but neither exon1 nor the lariat can be detected, as reverse transcription is blocked by the 2'-5' linkage at the branch point. Occasionally a reverse primer spanning the lariat branch structure (see the lariat paragraph of the qPCR section) taking advantage of the consensus sequences to design degenerate primers to all junctions was used to reverse prime. In this case signal from the lariats was seen in step 2 mutants and *dbr1Δ* mutant strains (see chapter 7).
C) mRNA and excised lariat after step two. The lariat cannot be detected, as it cannot be reverse transcribed and it is degraded. The mRNA can now be detected along with exon 2.
D) Table of which RNA species can be detected by each probe.

Equation 1: Normalised splicing ratio for microarray

$$Normalised\ splicing\ ratio = \left(\frac{\left(\frac{IntensityPT}{IntensityET}\right)}{\left(\frac{IntensityPC}{IntensityEC}\right)} \right)$$

Where
IntensityPT = Intensity of the Probe (5'SS, intron or mRNA) in the Test sample
IntensityET = Intensity of the Exon Probe in the Test sample
IntensityPC = Intensity of the Probe (5'SS, intron or mRNA) in the Control
IntensityEC = Intensity of the Exon Probe in the Control

1.8.2 RT-qPCR

Quantitative PCR (qPCR) or Real-Time PCR (RT-PCR) is a well-established technique. The use of the term "qPCR" is preferred over "Real Time PCR" to avoid confusion, as almost all qPCRs presented followed reverse transcription (RT), this will be made clear by the use of "RT-qPCR" in this work to indicate that qPCR was performed on cDNA reverse transcribed from RNA. This technique to monitor splicing in *S. cerevisiae* required optimisation as it was considered important to standardise all steps and conditions. Simple dye intercalation was used to monitor product accumulation. Initially this technique was developed to validate microarray results and borrowed much of their design from the microarrays but the system found general utility for many projects.

All PCR primers were designed using the same package (Oligo Perfect from ThermoFisher), using the same settings (Table 2), so all PCRs could be performed in the same standard conditions. PCR assays were validated for linearity over at least 3 orders of magnitude, for doubling each cycle within that linear range and specificity to their target. Any that did not meet these criteria were rejected.

Table 2: PCR design settings

Parameter	Value
PCR product size range	80-120 nt
Max Homopolymer length	4 nt
Max # of G or C allowed in last five 3' bases	4 Gs or Cs
Primer size range	20-35 nt
Tm Range	64.5 - 68 °C
Tm Optimum	65 °C
%GC	20-80 %

The ideal suite of PCR products to assay splicing and transcription is in Table 3 summarised in Figure 10. Note that none of the actual suites of PCR products for any gene conformed exactly to this ideal and alternatives were found where possible, also listed. The suite for the Ribo1 reporter (chapter 2) came closest to the ideal as the complete suite was validated. However, the lariat product was larger than preferred as the sequences differentiating it from the endogenous *ACT1* intron it was based on were further away from the BS than optimum. The *ACT1* transcript itself was the next most successful as only a 5'SS PCR could not be designed but a product over the branch site was deemed satisfactorily specific for pre-mRNA.

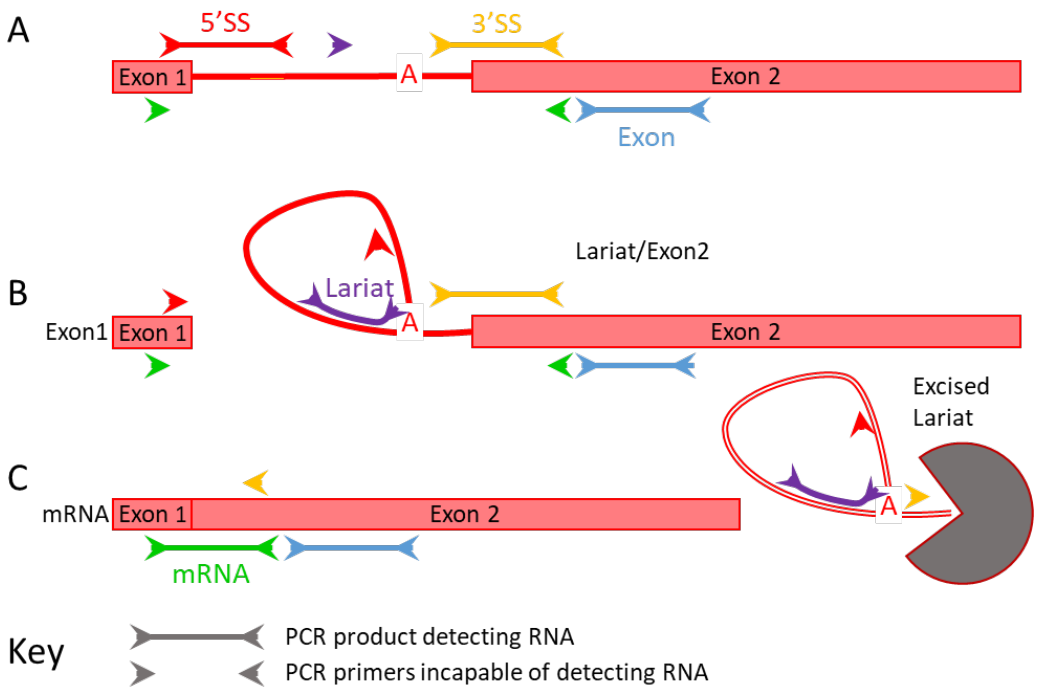
The PCR program used is presented in Table 4. This is the final version as performed in a Roche 480 Light cycler PCR machine (https://lifescience.roche.com/) with a 384 well block. The PCR programme did change but only to improve speed and throughput.

Table 3: PCR Program

Step	Temperature (ºC)	Time (sec)	Repeat
Hot start (enzyme activation)	94	120	-
Denaturation	94	10	40 times
Annealing / extension Fluorescence is read at the end of this step	60	15	
Melt curve Fluorescence is read at 0.5 °C intervals	60 to 94	As long as necessary to read the fluorescence	-

Table 4: PCR Assays

5'SS: the PCR spans the exon/intron junction at the 5'SS. This only detects pre-mRNA, lost after the first step of splicing. An alternative is for the PCR product to cross the branch point, this does not detect lariats, as the 2'-5' linkage blocks reverse transcription across this site.
3'SS: PCR spanning the intron/exon boundary at the 3'SS. This detects pre-mRNA and the lariat-exon2 species. A less specific alternative is a PCR product within the body of the intron, this needs a reverse transcription primer within the intron to detect lariats, and so also detects excised lariat.
Lariat: Used to detect the lariat-exon2 species and excised lariats, these cannot be distinguished, but comparison of results with the 3'SS PCR shows that the excised lariat cannot be a long-lived species; de-branching occurring soon after the second step. The 3' primer must be complementary to the region around the 5'SS, go across the BPA into the BS. The BPA with its 5'-2' phosphodiester bond in the lariat is read as a T in cDNAs that span the region (Vogel et al., 1997), so an A was placed at this point in the primer. The 3' end of the primer is to the region 5' of the BPA. Ideally 5-6 bases of the primer's 3' end is in the BS sequence. Too long a length in the BS results in the PCR also detecting the linear form of the intron; too little and no PCR product is produced. The reverse transcription must be performed with the reverse PCR primer as RT using primers 3’ of the branch point stalls at the branch point (Conklin et al., 2005). Success in designing these primers was very limited due to the generally AT rich regions in this area and the conserved nature of the BP sequence so the lariat PCR was limited to a small number of transcripts.
mRNA: PCR product crosses the exon/exon junction of the mature mRNA. Normally these PCRs are designed with one of the primers spanning the junction to improve specificity to mRNA versus gDNA. This was not done as mis-spliced as well as annotated splicing events were of interest and absolute specificity to only the annotated mRNA junction was not desired. All mRNA species are the product of splicing, so even miss-splicing is of interest to this study. It was hoped that mis-splicing events would be revealed as shoulders in the melt curve of the PCR product, allowing estimation of fidelity. In practice no indication of mis-splicing was seen, presumably because such aberrant events were swamped by the mass of annotated splice junctions. To avoid detection of pre-mRNA in this PCR, the annealing and extension steps were combined in a two-step PCR, using a temperature lower than optimum and extension step as short as possible to prevent completion of the longer product that would be produced from pre-mRNA, Table 4.
Exon: this PCR product was in exon 2 (exon 3 in the case of 2 intron genes). This PCR was used as a measure of transcript level. The PCR was designed to amplify a region as close to the intron as possible. No alternatives were necessary as there was no difficulty designing PCR primers to this region.
Controls: *ALG9* was frequently used as a housekeeping gene as its expression is remarkably stable (Teste et al., 2009). *scR1*, the RNAPIII generated signal recognition RNA (Felici et al., 1989), was used in experiments where it was suspected that RNAPII transcription could be disrupted. In practice it was not often used as it is such an abundant RNA species an extra dilution had to be performed to get the signal in the linear range. Latterly fixed *Schizosaccharomyces pombe* cells were spiked into samples and the *S. pombe* act1 gene was assayed by RT-qPCR, chapter 3.



D

	Pre-mRNA	Lariat /Exon2	Exon1	Excised Lariat	mRNA
5'SS	✓	×	×	×	×
3'SS	✓	✓	×	×	×
Lariat	×	✓	×	poorly	×
mRNA	×	×	×	×	✓
Exon	✓	✓	×	×	✓

Figure 10: RT-qPCR Assays of Splicing

The RNAs are as Figure 9. RT is performed by several primers, the most 3' being the reverse primer of the exon 2 primer set, one of the reverse primers for mRNA or 3'SS, whichever was most 3'. If a lariat PCR product had been designed, that would also be used to prime RT and finally the reverse primer of the 5'SS pair.

A) The 5'SS (red), 3'SS (yellow) and exon2 (blue) PCR reactions can all detect pre-mRNA

B) The 3'SS, exon2 and lariat (purple) PCRs can detect the lariat-exon2 product of the first step of splicing.

C) mRNA can be detected by the mRNA (green) and exon2 PCRs. Although it is possible that the excised lariat could be detected this is unlikely given the rapid debranching of the lariat.

D) The RNA species detected by each PCR product; using a combination of primer pairs, both catalytic steps of the splicing reaction can be monitored. The suite of ideal PCRs for a gene has much better resolving power than microarrays; only the isolated exon 1 cannot be detected.

The colours used for the PCRs are consistent throughout the thesis and some publications

Normalisation generally followed the strategy used with microarray data but modified for RT-qPCR, see Equation 2

Equation 2: Normalised splicing ratio for PCR

$$Normalised\ splicing\ ratio = \left(\frac{\left(2^{-(CtPT-CtET)}\right)}{\left(2^{-(CtPC-CtEC)}\right)}\right)$$

or

$$Normalised\ splicing\ ratio = 2^{-\left((CtPT-CtET)-(CtPC-CtEC)\right)}$$

Where
CtPT = Ct value for the 5'SS, 3'SS, Lariat or mRNA qPCR in the Test sample
CtET = Ct value for the exon qPCR in the Test sample
CtPC = Ct value for the 5'SS, 3'SS, Lariat or mRNA qPCR in the Control
CtEC = Ct value for the exon qPCR in the Control

Other normalisation strategies used the control PCRs mentioned in Table *3* (controls section). Spike in was only useful for short time courses, as cell growth leads to a general rise in signal for all other probes. *ALG9* and *scR1* normalisation gave similar results to exon normalisation, but the extra dilution required for *scR1* led to more variable results. The control, as for microarrays, was either wild type, unaltered (time point 0) or total RNA (steady state) for metabolic labelling studies. For ease of comparison, exon normalisation with a control sample was used for all PCRs reported in this thesis and is most used in the publications.

1.8.3 Thiolabelling RNA *in vivo*

This procedure is covered in detail in chapters 3 and 4. Chapter 3 is a methods paper. Briefly, 4-thiouracil is added to a culture of growing cells. This incorporates into RNA as it is transcribed *in vivo*, without disturbing growth during the short time courses used. RNA is purified and a biotin moiety is attached to the thio group. Streptavidin bound to magnetic beads captures this newly synthesised RNA (nsRNA). The S-S bond reduced to release the nsRNA. RNA from labelling experiments down to 10 seconds is detectable but 15 seconds is less error prone.

1.8.4 AID Degron

This procedure is covered in detail in chapters 6 and 7. Chapter 6 describes the method. A means of action of the plant hormone auxin is the degradation of target proteins using the ubiquitin system. This system has been recapitulated in *S. cerevisiae* so a suitably tagged protein can be very quickly depleted with little other impact on the cell.

1.9 Research publications

All refereed research publications on which I am an author are listed below. Papers published in October 2009 and onwards are eligible for inclusion in this thesis (see also Table 5). All those chosen are concerned with RNA processing (splicing in particular), and how this interacts with transcription. As with any discussion on these two processes degradation and recycling has to be considered. Two chapters are detailed descriptions of the methods used, with particular focus on transcription, RNA processing and finally recycling (Figure 11 and Table 5).

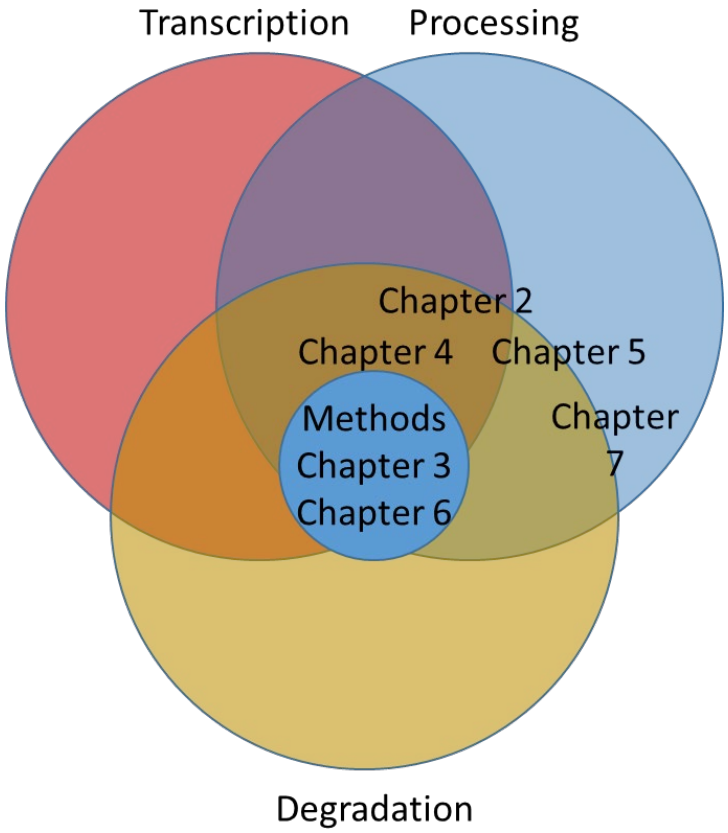


Figure 11: Chapter and paper topics
Chapters 2 and particularly 4 are general RNA life cycle papers. All other chapters focus on narower aspects of RNA processing.
Chapter 5 deals almost exclusively with pre-mRNA splicing, but degradation is also considered. Chapter 7 covers similar topics, but with an emphasis on errors in splicing leading to a recycling defect resulting in degradation. Chapters 3 and 6 are on methods, the method in chapter 6 is of general utillity, but showing a splicing application, while Chapter 3's method is RNA focussed.

Table 5: Publications discussed in each chapter

Chapter	Publication(s)
2	Alexander, R. D., Barrass, J.D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M.-C., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., Dichtl, B., Kufel, J., Bertrand, E., Beggs, J.D., 2010. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. RNA 16, 2570–2580.
3	Barrass, J.D., Beggs, J.D., 2019. Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE (Journal of Visualized Experiments) e59952.
4	Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., Granneman, S., 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biology 16, 282.
5	Kafasla, P., Barrass, J.D., Thompson, E., Fromont-Racine, M., Jacquier, A., Beggs, J.D., Lewis, J., 2009. Interaction of yeast eIF4G with spliceosome components. RNA Biol 6, 563–574.
6	Mendoza-Ochoa, G.I., Barrass, J.D., Terlouw, B.R., Maudlin, I.E., Lucas, S. de, Sani, E., Aslanzadeh, V., Reid, J.A.E., Beggs, J.D., 2018. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. Yeast 36 (1) 75-81. Barrass, J.D., Mendoza-Ochoa, G.I., Maudlin, I.E., Sani, E., Beggs, J.D., 2019. Tuning Degradation to Achieve Specific and Efficient Protein Depletion. JoVE (Journal of Visualized Experiments) e59874.
7	Mendoza-Ochoa, G.I., Barrass, J.D., Maudlin, I.E., Beggs, J.D., 2019. Blocking late stages of splicing quickly limits pre-spliceosome assembly in vivo. RNA Biol 16, 1775–1784.

1.10 Complete List of Publications

Publications submitted in this thesis are bold

Kennedy, P.G.E., Barrass, J.D., Graham, D.I., Clemens, G.B., 1990. Studies on the pathogenesis of neurological diseases associated with Varicella-Zoster Virus. Neuropathology and Applied Neurobiology 16, 305–316.

Carver, A., Wright, G., Cottom, D., Cooper, J., Dalrymple, M., Temperley, S., Udell, M., Reeves, D., Percy, J., Scott, A., Barrass, D., Gibson, Y., Jeffrey, Y., Samuel, C., Colman, A., Garner, I., 1992. Expression of human α1 antitrypsin in transgenic sheep. Cytotechnology 9, 77–84.

Carver, A.S., Dalrymple, M.A., Wright, G., Cottom, D.S., Reeves, D.B., Gibson, Y.H., Keenan, J.L., Barrass, J.D., Scott, A.R., Colman, A., Garner, I., 1993. Transgenic Livestock as Bioreactors: Stable Expression of Human Alpha-1-Antitrypsin by a Flock of Sheep. Nat Biotech 11, 1263–1270.

Barrass, J.D., Beggs, J.D., 2003. Splicing goes global. Trends in Genetics 19, 295–298.

Boon, K.-L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J.D., Droop, A.P., Dez, C., Beggs, J.D., 2006. Yeast Ntr1/Spp382 Mediates Prp43 Function in Postspliceosomes. Mol. Cell. Biol. 26, 6016–6023.

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., Tollervey, D., 2006. Microarray detection of novel nuclear RNA substrates for the exosome. Yeast 23, 439–454.

Boon, K.-L., Grainger, R.J., Ehsani, P., Barrass, J.D., Auchynnikava, T., Inglehearn, C.F., Beggs, J.D., 2007. prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. Nat Struct Mol Biol 14, 1077–1083.

Kershaw, C.J., Barrass, J.D., Beggs, J.D., O'Keefe, R.T., 2009. Mutations in the U5 snRNA result in altered splicing of subsets of pre-mRNAs and reduced stability of Prp8. RNA 15, 1292–1304.

Kafasla, P., Barrass, J.D., Thompson, E., Fromont-Racine, M., Jacquier, A., Beggs, J.D., Lewis, J., 2009. Interaction of yeast eIF4G with spliceosome components. RNA Biol 6, 563–574.

Grainger, R.J., Barrass, J.D., Jacquier, A., Rain, J.-C., Beggs, J.D., 2009. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. RNA 15, 2161–2173.

Alexander, R. D., Barrass, J.D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M.-C., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., Dichtl, B.,

Kufel, J., Bertrand, E., Beggs, J.D., 2010. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. RNA 16, 2570–2580.

Alexander, Ross D, Innocente, S.A., Barrass, J.D., Beggs, J.D., 2010. Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell 40, 582–593.

Swiatkowska, A., Wlotzka, W., Tuck, A., Barrass, J.D., Beggs, J.D., Tollervey, D., 2012. Kinetic analysis of pre-ribosome structure in vivo. RNA 18, 2187–2200.

Chathoth, K.T., Barrass, J.D., Webb, S., Beggs, J.D., 2014. A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. Molecular Cell 53, 779–790.

Cordin, O., Hahn, D., Alexander, R., Gautam, A., Saveanu, C., Barrass, J.D., Beggs, J.D., 2014. Brr2p carboxy-terminal Sec63 domain modulates Prp16 splicing RNA helicase. Nucleic Acids Res 42, 13897–13910.

Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., Granneman, S., 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biology 16, 282.

Gautam, A., Grainger, R.J., Vilardell, J., Barrass, J.D., Beggs, J.D., 2015. Cwc21p promotes the second step conformation of the spliceosome and modulates 3′ splice site selection. Nucleic Acids Res 43, 3309–3317.

Mendoza-Ochoa, G.I., Barrass, J.D., Terlouw, B.R., Maudlin, I.E., Lucas, S. de, Sani, E., Aslanzadeh, V., Reid, J.A.E., Beggs, J.D., 2018. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. Yeast 36 (1) 75-81.

Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., Jensen, T.H., 2018. A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA. Cell Reports 24, 2457-2467.

Barrass, J.D., Mendoza-Ochoa, G.I., Maudlin, I.E., Sani, E., Beggs, J.D., 2019. Tuning Degradation to Achieve Specific and Efficient Protein Depletion. JoVE (Journal of Visualized Experiments) e59874.

Barrass, J.D., Beggs, J.D., 2019. Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE (Journal of Visualized Experiments) e59952.

Mendoza-Ochoa, G.I., Barrass, J.D., Maudlin, I.E., Beggs, J.D., 2019. Blocking late stages of splicing quickly limits pre-spliceosome assembly in vivo. RNA Biol 16, 1775–1784.

1.11 References

Aebi, M., Hornig, H., and Weissmann, C. (1987). 5′ cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5′ splice region, not by the conserved 5′ GU. Cell *50*, 237–246.

Aguilera, A., and García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Mol. Cell *46*, 115–124.

Aitken, S., Alexander, R.D., and Beggs, J.D. (2011). Modelling Reveals Kinetic Advantages of Co-Transcriptional Splicing. PLoS Comput Biol *7*, e1002215.

Aitken, S., Alexander, R.D., and Beggs, J.D. (2013). A rule-based kinetic model of RNA polymerase II C-terminal domain phosphorylation. J. R. Soc. Interface *10*, 20130438.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010a). Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell *40*, 582–593.

Alexander, R.D., Barrass, J.D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M.-C., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., et al. (2010b). RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in Saccharomyces cerevisiae. RNA *16*, 2570–2580.

Alpert, T., Straube, K., Carrillo Oesterreich, F., Herzel, L., and Neugebauer, K.M. (2020). Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. Cell Rep. *33*, 108324.

Ares, M.Jr., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA *5*, 1138–1139.

Aslanzadeh, V., Huang, Y., Sanguinetti, G., and Beggs, J.D. (2018). Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. Genome Res. *28*, 203–213.

Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., and Cramer, P. (2014). Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. Mol. Cell *55*, 745–757.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A Ratchet Mechanism of Transcription Elongation and Its Control. Cell *120*, 183–193.

Barrass, J.D., and Beggs, J.D. (2003). Splicing goes global. Trends Genet *19*, 295–298.

Barrass, J.D., and Beggs, J.D. (2019). Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE J. Vis. Exp. *150*, e59952.

Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. *16*, 282.

Barrass, J.D., Mendoza-Ochoa, G.I., Maudlin, I.E., Sani, E., and Beggs, J.D. (2019). Tuning degradation to achieve specific and efficient protein depletion. JoVE J. Vis. Exp. J Vis Exp. *142,* e59874.

Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318–323.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev *2*, 754–765.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Mol. Cell *40*, 571–581.

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell*165,* 372-38.

Carrocci, T.J., and Neugebauer, K.M. (2020). Pre-mRNA Splicing in the Nuclear Landscape. Cold Spring Harb. Symp. Quant. Biol. *84,* 11-20.

Chanarat, S., and Svasti, J. (2020). Stress-induced upregulation of the ubiquitin-relative Hub1 modulates pre-mRNA splicing and facilitates cadmium tolerance in Saccharomyces cerevisiae. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1867,* 118565.

Chang, T.-H., Tung, L., Yeh, F.-L., Chen, J.-H., and Chang, S.-L. (2013). Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829* 764-74

Chapman, K.B., and Boeke, J.D. (1991). Isolation and Characterization of the Gene Encoding Yeast Debranching Enzyme. Cell *65*, 483–492.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. Mol. Cell *53*, 779–790.

Chen, H.-C., and Cheng, S.-C. (2012). Functional roles of protein splicing factors. Biosci. Rep. *32*, 345–359.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, 700–705.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev *11*, 3319–3326.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422.

Conklin, J.F., Goldman, A., and Lopez, A.J. (2005). Stabilization and analysis of intron lariats in vivo. Methods *37*, 368–375.

Conti, L.D., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. WIREs RNA *4*, 49–60.

Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. RNA Biol. *10*, 83–95.

Courchaine, E.M., Lu, A., and Neugebauer, K.M. (2016). Droplet organelles? EMBO J. *35*, 1603–1612.

Cramer, P. (2002). Multisubunit RNA polymerases. Curr. Opin. Struct. Biol. *12*, 89–97.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

Cvitkovic, I., and Jurica, M.S. (2013). Spliceosome database: a tool for tracking components of the spliceosome. Nucleic Acids Res. *41*, 132-141.

Davis, C.A., Grate, L., Spingola, M., and Ares Jr, M. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res. *28*, 1700–1706.

De, I., Schmitzová, J., and Pena, V. (2016). The organization and contribution of helicases to RNA splicing. Wiley Interdiscip. Rev. RNA *7*, 259-74..

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

Dlakić, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA *17*, 799–808.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell *77*, 985-998.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. Mol. Cell *36*, 593–608.

Felici, F., Cesareni, G., and Hughes, J.M. (1989). The most abundant small cytoplasmic RNA of Saccharomyces cerevisiae has an important function required for normal cell growth. Mol. Cell. Biol. *9*, 3260–3268.

Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nat. Struct. Mol. Biol. *24*, 791–799.

Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.-C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. Nature *542*, 377–380.

García-Martínez, J., Aranda, A., and Pérez-Ortín, J.E. (2004). Genomic Run-On Evaluates Transcription Rates for All Yeast Genes and Identifies Gene Regulatory Mechanisms. Mol. Cell *15*, 303–313.

Gibbs, E.B., Lu, F., Portz, B., Fisher, M.J., Medellin, B.P., Laremore, T.N., Zhang, Y.J., Gilmour, D.S., and Showalter, S.A. (2017). Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. Nat. Commun. *8*, 1–11.

Giorgi, C., Fatica, A., Nagel, R., and Bozzoni, I. (2001). Release of U18 snoRNA from its host intron requires interaction of Nop1p with the Rnt1p endonuclease. EMBO J. *20*, 6856–6865.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Mol. Cell *19*, 53–63.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: At the heart of the spliceosome. RNA *11*, 533–557.

Grünberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem. Sci. *38*, 603–611.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive degradation of RNA precursors by the exosome in wild type cells. Mol. Cell *48*, 409–421.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. *19*, 621–637.

Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. Science *349*, 1191–1198.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.

Hartman, E., Wang, Z., Zhang, Q., Roy, K., Chanfreau, G., and Feigon, J. (2013). Intrinsic dynamics of an extended hydrophobic core in the S. cerevisiae RNase III dsRBD contributes to recognition of specific RNA binding sites. J. Mol. Biol. *425*, 546–562.

Hilliker, A.K., Mefford, M.A., and Staley, J.P. (2007). U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev. *21*, 821–834.

Hooks, K.B., Naseeb, S., Parker, S., Griffiths-Jones, S., and Delneri, D. (2016). Novel Intronic RNA Structures Contribute to Maintenance of Phenotype in Saccharomyces cerevisiae. Genetics *203*, 1469–1481.

Huranová, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., Shav-Tal, Y., Neugebauer, K.M., and Staněk, D. (2010). The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cellsSpliceosome assembly in the cell nucleus. J. Cell Biol. *191*, 75–86.

Husain, A., Begum, N.A., Taniguchi, T., Taniguchi, H., Kobayashi, M., and Honjo, T. (2016). Chromatin remodeller SMARCA4 recruits topoisomerase 1

and suppresses transcription-associated genomic instability. Nat. Commun. *7*, 10549.

Imashimizu, M., Shimamoto, N., Oshima, T., and Kashlev, M. (2014). Transcription elongation. Heterogeneous tracking of RNA polymerase and its biological implications. Transcription.*5*, e28285.

Juneau, K., Nislow, C., and Davis, R.W. (2009). Alternative Splicing of PTC7 in Saccharomyces cerevisiae Determines Protein Localization. Genetics *183*, 185–194.

Koodathingal, P., and Staley, J.P. (2013). Splicing fidelity. RNA Biol. *10*, 1073–1079.

Koodathingal, P., Novak, T., Piccirilli, J.A., and Staley, J.P. (2010). The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5′ Splice Site Cleavage during Pre-mRNA Splicing. Mol. Cell *39*, 385–395.

Kotovic, K.M., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional Recruitment of the U1 snRNP to Intron-Containing Genes in Yeast. Mol. Cell. Biol. *23*, 5768–5779.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., and Davis, R.W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U. S. A. *94*, 13057–13062.

Le Hir, H., Nott, A., and Moore, M.J. (2003). How introns influence and enhance eukaryotic gene expression. Trends Biochem. Sci. *28*, 215–220.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*, 220–224.

Leung, C.S., Douglass, S.M., Morselli, M., Obusan, M.B., Pavlyukov, M.S., Pellegrini, M., and Johnson, T.L. (2019). H3K36 Methylation and the Chromodomain Protein Eaf3 Are Required for Proper Cotranscriptional Spliceosome Assembly. Cell Rep. *27*, 3760-3769.

Lewis, J.D., Gorlich, D., and Mattaj, I.W. (1996). A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucl.Acids.Res. *24*, 3332–3336.

Liang, W.-W., and Cheng, S.-C. (2015). A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. Genes Dev. *29*, 81–93.

Libri, D. (2010). Nuclear Poly(A)-Binding Proteins and Nuclear Degradation: Take the mRNA and Run? Mol. Cell *37*, 3–5.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. Mol. Cell. Biol. *29*, 4852–4863.

Lopez, P.J., and Séraphin, B. (2000). YIDB: the Yeast Intron DataBase. Nucleic Acids Res. *28*, 85–86.

Luo, W., and Bentley, D. (2004). A Ribonucleolytic Rat Torpedoes RNA Polymerase II. Cell *119*, 911–914.

Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I., and Yu, Y.-T. (2005). Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. EMBO J. *24*, 2403–2413.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997). 5’ -Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306–3318.

McGrail, J.C., and O’Keefe, R.T. (2008). The U1, U2 and U5 snRNAs crosslink to the 5′ exon during yeast pre-mRNA splicing. Nucleic Acids Res. *36*, 814–825.

Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E., and Vilardell, J. (2011). Deciphering 3′ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. Mol. Cell *43*, 1033–1039.

Milligan, L., Huynh-Thu, V.A., Delan-Forino, C., Tuck, A., Petfalski, E., Lombraña, R., Sanguinetti, G., Kudla, G., and Tollervey, D. (2016). Strand-specific, high-resolution mapping of modified RNA polymerase II. Mol. Syst. Biol. *12*, 874.

Milligan, L., Sayou, C., Tuck, A., Auchynnikava, T., Reid, J.E., Alexander, R., Alves, F. de L., Allshire, R., Spanos, C., Rappsilber, J., et al. (2017). RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. ELife *Oct 13, e27082*.

Neugebauer, K.M. (2019). Nascent RNA and the Coordination of Splicing with Transcription. Cold Spring Harb. Perspect. Biol. *11* a032227.

Neves, L.T., Douglass, S., Spreafico, R., Venkataramanan, S., Kress, T.L., and Johnson, T.L. (2017). The histone variant H2A.Z promotes efficient cotranscriptional splicing in S. cerevisiae. Genes Dev. *31*, 702–717.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

Ooi, S.L., Samarsky, D.A., Fournier, M.J., and Boeke, J.D. (1998). Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: Intron length effects and activity of a precursor snoRNA. RNA *4*, 1096–1110.

Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., Cecez, B., Gervais-Bird, J., Koh, C.-S., Brunelle, D., et al. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. Mol. Biol. Cell *19*, 1932–1941.

Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.-F., Wellinger, R.J., Chabot, B., and Abou Elela, S. (2011). Introns within Ribosomal Protein Genes Regulate the Production and Function of Yeast Ribosomes. Cell *147*, 320–331.

Parenteau, J., Maignon, L., Berthoumieux, M., Catala, M., Gagnon, V., and Elela, S.A. (2019). Introns are mediators of cell response to starvation. Nature *565*, 612.

Parker, R. (2012). RNA Degradation in Saccharomyces cerevisae. Genetics *191*, 671–702.

Peck, S.A., Hughes, K.D., Victorino, J.F., and Mosley, A.L. (2019). Writing a wrong: Coupled RNA polymerase II transcription and RNA quality control. WIREs RNA *10*, e1529.

Perriman, R.J., and Ares, M. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. Genes Dev. *21*, 811–820.

Piekna-Przybylska, D., Decatur, W.A., and Fournier, M.J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. RNA *13*, 305–312.

Plaschka, C., Newman, A.J., and Nagai, K. (2019). Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb. Perspect. Biol. May 1;*11* a032391.

Plass, M., Codony-Servat, C., Ferreira, P.G., Vilardell, J., and Eyras, E. (2012). RNA secondary structure mediates alternative 3′ss selection in Saccharomyces cerevisiae. RNA *18*, 1103–1115.

Porrua, O., and Libri, D. (2013). RNA quality control in the nucleus: The Angels' share of RNA. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 604–611.

Proudfoot, N. (2000). Connecting transcription to messenger RNA processing. Trends Biochem. Sci. *25*, 290–293.

Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. Mol. Cell S1097-2765(20)30937-0 (Epub ahead of print).

Rose, A.B. (2019). Introns as Gene Regulators: A Brick on the Accelerator. Front. Genet. Feb *7*, 672.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. Genes Dev. *20*, 1050–1056.

Safina, A., Cheney, P., Pal, M., Brodsky, L., Ivanov, A., Kirsanov, K., Lesovaya, E., Naberezhnov, D., Nesher, E., Koman, I., et al. (2017). FACT is a sensor of DNA torsional stress in eukaryotic cells. Nucleic Acids Res. *45*, 1925–1945.

Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J. Mol. Biol. *428*, 2623–2635.

Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. Nat. Rev. Genet. *19*, 518–529.

Schreiber, K., Csaba, G., Haslbeck, M., and Zimmer, R. (2015). Alternative Splicing in Next Generation Sequencing Data of Saccharomyces cerevisiae. PLOS ONE *10*, e0140487.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol. Cell *61*, 305–314.

Schwartz, S., and Ast, G. (2010). Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. EMBO J. *29*, 1629–1636.

Semlow, D.R., and Staley, J.P. (2012). Staying on message: ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci. *37*, 263–273.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell *59*, 349–358.

Shandilya, J., and Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1819*, 391–400.

Shao, W., Ding, Z., Zheng, Z.-Z., Shen, J.-J., Shen, Y.-X., Pu, J., Fan, Y.-J., Query, C.C., and Xu, Y.-Z. (2020). Prp5−Spt8/Spt3 interaction mediates a reciprocal coupling between splicing and transcription. Nucleic Acids Res. *48*, 5799–5813.

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. *18*, 655–670.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nat. Struct. Mol. Biol. *19*, 594–601.

Sohrabi-Jahromi, S., Hofmann, K.B., Boltendahl, A., Roth, C., Gressel, S., Baejen, C., Soeding, J., and Cramer, P. (2019). Transcriptome maps of general eukaryotic RNA degradation factors. ELife *8*, e47040.

Sorenson, M.R., Jha, D.K., Ucles, S.A., Flood, D.M., Strahl, B.D., Stevens, S.W., and Kress, T.L. (2016). Histone H3K36 methylation regulates pre-mRNA splicing in Saccharomyces cerevisiae. RNA Biol. *13*, 412–426.

Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. Cold Spring Harb. Perspect. Biol. *3* a000646

Spingola, M., Grate, L., Haussler, D., and Ares, M.Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA *5*, 221–234.

Suh, H., Hazelbaker, D.Z., Soares, L.M., and Buratowski, S. (2013). The C-Terminal Domain of Rpb1 Functions on Other RNA Polymerase II Subunits. Mol. Cell *51*, 850-8.

Suh, H., Ficarro, S.B., Kang, U.-B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol. Cell *61*, 297–304.

Svetlov, V., and Nudler, E. (2013). Basic mechanism of transcription by RNA polymerase II. Biochim. Biophys. Acta *1829*, 20–28.

Tang, Q., Rodriguez-Santiago, S., Wang, J., Pu, J., Yuste, A., Gupta, V., Moldón, A., Xu, Y.-Z., and Query, C.C. (2016). SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev. *30*, 2710–2723.

Tardiff, D.F., Lacadie, S.A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol. Cell *24*, 917–929.

Teste, M.-A., Duquenne, M., Francois, J., and Parrou, J.-L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. BMC Mol. Biol. *10*, 99.

Timmers, H.T.M., and Tora, L. (2018). Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. Mol. Cell *72*, 10–17.

Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., and Jensen, T.H. (2018). A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA. Cell Rep. *24*, 2457-2467.

Ubukata, T., Shimizu, T., Adachi, N., Sekimizu, K., and Nakanishi, T. (2003). Cleavage, but Not Read-through, Stimulation Activity Is Responsible for Three Biologic Functions of Transcription Elongation Factor S-II. J. Biol. Chem. *278*, 8580–8585.

Vilardell, J., and Warner, J.R. (1997). Ribosomal protein L32 of Saccharomyces cerevisiae influences both the splicing of its own transcript and the processing of rRNA. Mol Cell Bio *17*, 1959–1965.

Vogel, J., Hess, W.R., and Börner, T. (1997). Precise branch point mapping and quantification of splicing intermediates. Nucleic Acids Res. *25*, 2030–2031.

Volanakis, A., Passoni, M., Hector, R.D., Shah, S., Kilchert, C., Granneman, S., and Vasiljeva, L. (2013). Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. Genes Dev. *27*, 2025–2038.

Wallace, E.W.J., and Beggs, J.D. (2017). Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. RNA *23*, 601–610.

Wan, R., Bai, R., Yan, C., Lei, J., and Shi, Y. (2019). Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. Cell *177*, 339-351.

Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., Newman, A.J., and Nagai, K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3′–splice site selection. Science *358*, 1283–1288.

Will, C.L., and Luhrmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. *3,* a003707.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. Cell *121*, 725–737.

Xu, Y.-Z., and Query, C.C. (2007). Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. Mol. Cell *28*, 838–849.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. Science *355*, 149–155.

Yang, F., Wang, X.-Y., Zhang, Z.-M., Pu, J., Fan, Y.-J., Zhou, J., Query, C.C., and Xu, Y.-Z. (2013). Splicing proofreading at 5′ splice sites by ATPase Prp28p. Nucleic Acids Res. *41*, 4660–4670.

Yogesha, S.D., Mayfield, J.E., and Zhang, Y. (2014). Cross-Talk of Phosphorylation and Prolyl Isomerization of the C-terminal Domain of RNA Polymerase II. Molecules *19*, 1481–1511.

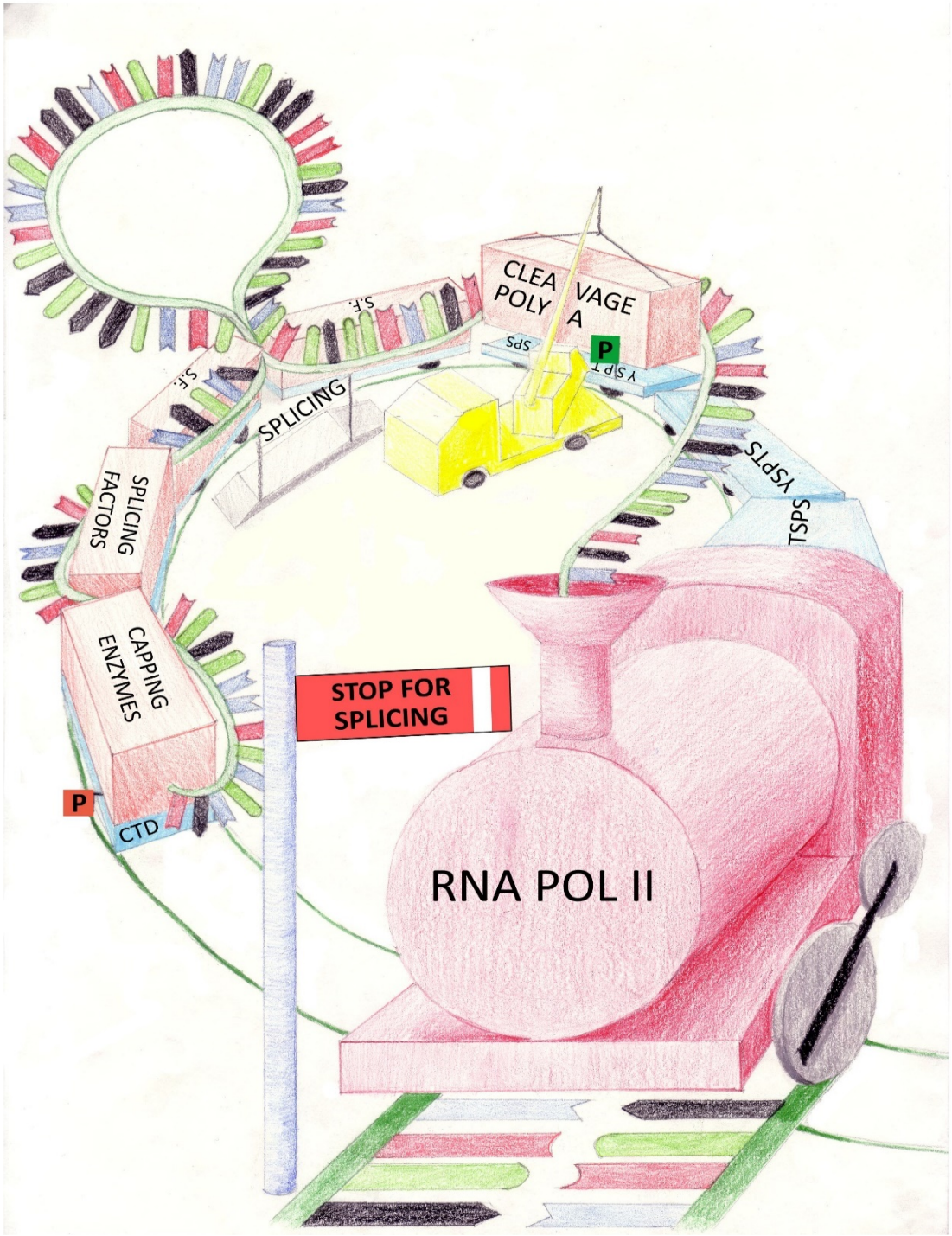
Yurko, N.M., and Manley, J.L. (2018). The RNA polymerase II CTD “orphan” residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. Transcription *9*, 30–40.

Yurko, N., Liu, X., Yamazaki, T., Hoque, M., Tian, B., and Manley, J.L. (2017). MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. Mol. Cell *68*, 913-925.

Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. Nat. Struct. Mol. Biol. *23*, 771–777.

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. Cell *169*, 918-929.

2 RiboSys: Towards a Kinetic Understanding of Transcription and RNA Processing



Cover image for Alexander, Ross D, Innocente, S.A., Barrass, J.D., Beggs, J.D., 2010. **Splicing-dependent RNA polymerase pausing in yeast.** Mol. Cell 40, 582–593. Drawn by my 15 year old Daughter, Chrissie, and used as the front cover for that issue of Molecular Cell. RNAPII is a train on DNA tracks, RNA being produced. The carriages are the CTD repeats with factors loaded

2.1 Research Articles

This chapter is based on the peer-reviewed publication: Alexander, R. D., Barrass, J.D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M.-C., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., Dichtl, B., Kufel, J., Bertrand, E., Beggs, J.D., 2010. RiboSys, a high-resolution, quantitative approach to measure the *in vivo* kinetics of pre-mRNA splicing and 3'-end processing in *Saccharomyces cerevisiae*. RNA 16, 2570–2580.

Figures from this publication will be referred to as P.Figure followed by the number.

The related publication: Alexander, Ross D, Innocente, S.A., Barrass, J.D., Beggs, J.D., 2010. Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell 40, 582–593 is not part of this submission. Although I am an author, my contribution was in method development rather than results. This publication does, however, use the techniques and some data from the submitted publication and follows on from it.

2.2 Aim

The aim of the RiboSys project was, in the words of the grant proposal, "to model RNA metabolism in yeast to aid understanding of these complex cellular pathways. In order to develop kinetic models, we propose to quantify RNA precursors and directly determine their rates of production as well as their processing and degradation through the various post-transcriptional pathways".

This research article examines the production and processing of reporter transcripts, especially their splicing and 3' end formation. The explicit aim of this study was to discover levels of expression, rates of processing and rates of degradation (expressed as half-lives). A paper describing mathematical modelling of the data was published separately (Aitken et al., 2010)

2.3 Experiments

All experiments were performed using RNA transcribed from the Ribo1 system of artificial reporters in the yeast *S. cerevisiae*. RT-qPCR and *in situ* hybridisation were used to monitor transcription, processing and degradation. Two reporter systems were developed; one that could be induced (Ribo1 tetON) and one repressed (Ribo1 tetOFF)

2.3.1 Experimental system

The crucial feature of the artificial reporter system was that its transcription could be induced from a low basal level or repressed from a high level very rapidly. All without perturbing the cell, other than the distribution of transcription and processing components inevitable on inducing or repressing the reporter's transcription. Common yeast inducible promoters were considered but rejected as their induction has considerable impact on the cell's metabolism. The tetracycline-inducible (tet) system was used as tetracycline (or more accurately its analogue doxycycline) does not impact the cell metabolism (see references in the paper). Figure 1 shows the version of the reporter where doxycycline induces expression and its absence represses expression, this is Ribo1 tetON. Ribo1 tetOFF is the second reporter and its regulation is reversed; doxycycline represses expression; strongly expressed in its absence.

Ribo1 is a single intron-containing gene based on *PGK1* gene with the intron from *ACT1* (shown in more detail in P.Figure 1). Several non-yeast sequences were added; of particular relevance is a λ phage BoxB sequence towards the 5' end of the intron. These additional sequences and fusion sites were the only places distinguishing the reporter from the endogenous *ACT1* or *PGK1* genes so they were used as unique sites for *in situ* hybridisation probes or RT-qPCR primer binding sites.

Three other reporters were made, based on the Ribo1 tetON plasmid, with the 5'SS (5'SS Ribo1) or 3'SS (3'SS Ribo1) mutated to be non-functional and one without an intron (IL Ribo1). The Ribo1 tetON system, and its derivatives, was

extensively used for expression and processing experiments, Ribo1 tetOFF for degradation, half-life timing studies.

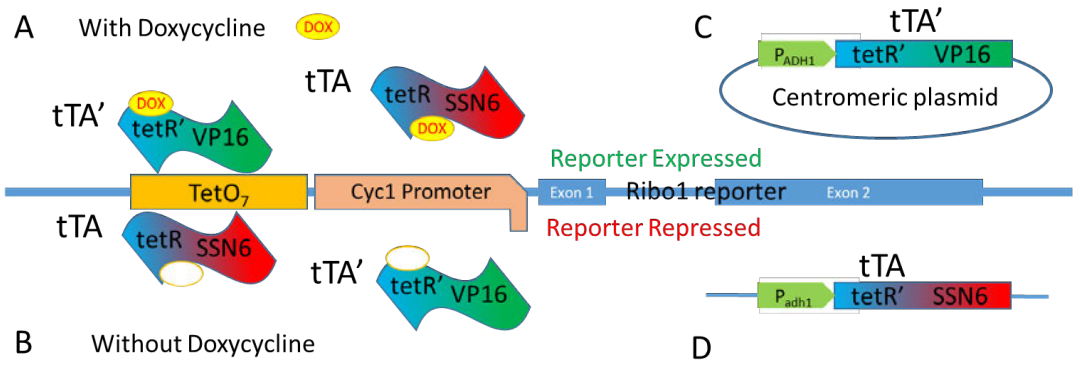


Figure 1: The Ribo system
The Ribo system is a single intron reporter gene with a proximal promoter region from CYC1. Its expression is controlled by the 7 tet repressor binding sites ($TetO_7$) at the 5' end of the promoter and the tet transactivator proteins (tTA and tTA') that can bind to it. Shown here is the **Ribo1 tetON** construct.
A) When doxycycline (DOX) is added the tet repressor (tetR) of the tTA binds doxycycline then dissociates from the $TetO_7$. The modified tet repressor (tetR') of the tTA' also binds doxycycline and replaces tTA on $TetO_7$. The herpes virus gene activator VP16 fused to the tetR' activates the promoter, stimulating transcription of the Ribo reporter.
B) The reverse happens in the absence of doxycycline, the Ribo reporter is repressed by the stimulatory tetR' being displaced from the $TetO_7$ sites by tetR. tTA is fused to the repressive transcription factor Ssn6p (now known as Cyc8) in tetR.
C) To achieve sufficient production of tTA' it was placed in a centromeric plasmid under the control of the *S. cerevisiae ADH1* promoter.
D) tTA was integrated into the genome and expressed from the *S. pombe adh1* promoter.
The **Ribo1 tetOFF** is functionally similar, except the tTA fusions are reversed, tetR' with *SSN6* and tetR with VP16, resulting in reversing the regulation of reporter expression. Additionally, both the tTAs were expressed from the genome.

2.3.2 Quantification of Transcripts in Populations of Cells

Due to the known inadequacy of housekeeping genes for estimation of copy number (Thellin et al., 1999), a new system was needed to estimate reporter transcript numbers per cell (P.Figure 2). Cell number estimation in yeast proved difficult, mainly due to the cells clumping in culture (Ribo containing

strains seemed particularly prone to this), hence the NucleoCounter (ChemoMetec) was used to count discrete nuclei. To calculate the RNA content of cells, losses of RNA at each stage of extraction and quantitation were estimated. Losses during extraction were measured by spiking the sample with a known amount of *in vitro* transcribed RNA at the very first stage of extraction. The RNA spike was added to the lysis mixture, a solution containing chaotropic agents and phenol, stored at 4 ºC, so the RNA was protected from degradation. This mixture was used for every extraction, so the same spike level was used for many samples and experiments. Losses during extraction and efficiency of reverse transcription were assayed against recovery of the spike. To control for cell lysis efficiency, DNA was also prepared from the lysed cells (the procedure allowed both to be prepared from the same cell pellet) and compared to the DNA content per nucleus. This is estimated from the number of nuclei and FACs analysis of genome content of log phase growing yeast. Comparing the DNA recovered with the calculated DNA content allows an estimation of lysis efficiency.

2.3.3 De-repression and Induction Kinetics

Transcription and splicing were assayed using the PCR system outlined in the Introduction. The full set of PCRs products to detect 5'SS, 3'SS, mRNA, Exon and Lariat were designed. The lariat PCR product was longer than optimal to distinguish the Ribo1 and *ACT1* lariats, as Ribo1's intron was that of *ACT1* except for the BoxB sequence added near the 5' end. Other PCRs achieved their specificity by crossing the *PGK1*/*ACT1* fusion junctions.

Production of transcripts from both the Ribo1 tetON and tetOFF reporters were assayed (P.Figure 5). To induce the Ribo1 tetOFF reporter the growth medium had to be changed by centrifugation, washing the cell pellet and resuspension in doxycycline-free medium. This perturbs the cells and is unlikely to completely remove the doxycycline, probably explaining the slower and less powerful induction seen with the tetOFF reporter (A-C compared to E-F). The results using this strain will not be discussed here, other than to point out that they follow the same trends as the Ribo1 tetON reporter.

The results shown in P.Figure 5, of which a portion is reproduced, modified, here (Figure 2), illustrates several kinetic parameters of the tetON series of reporters. The induced reporter RNA is first seen in the sample taken 5 minutes after doxycycline addition (D - F). This is an indication of the time taken to initiate transcription upon addition of the inducer.

Steady state levels are when production and removal are in balance. Production of RNA is by transcription and some processing events (e.g. splicing creating the lariat and mRNA). Removal can also be by processing (e.g. splicing of pre-mRNA), and degradation.

For pre-mRNA (D & E) <7.5 minutes is needed for transcription to match both splicing (Ribo1) and degradation (5'SS Ribo1). The levels of pre-mRNA are similar for both reporters (D & E), as is the time scale, implying that there is a certain time for the first step of splicing to occur (D) otherwise surveillance reduces accumulation (E). The 5'SS Ribo1 reporter has no 5'SS so U1 cannot bind, the spliceosome cannot itself be the surveillance mechanism. Rather retention of the intron must be the signal, possibly via NMD, and could signal to stop transcription. mRNA (D), rises at the same time as pre-mRNA in Ribo1 but only after 12.5 minutes does degradation equal production by transcription and splicing. The level of mRNA is much higher than the pre-mRNA, presumably this is due to mRNA’s higher stability and continuous production from pre-mRNA by splicing.

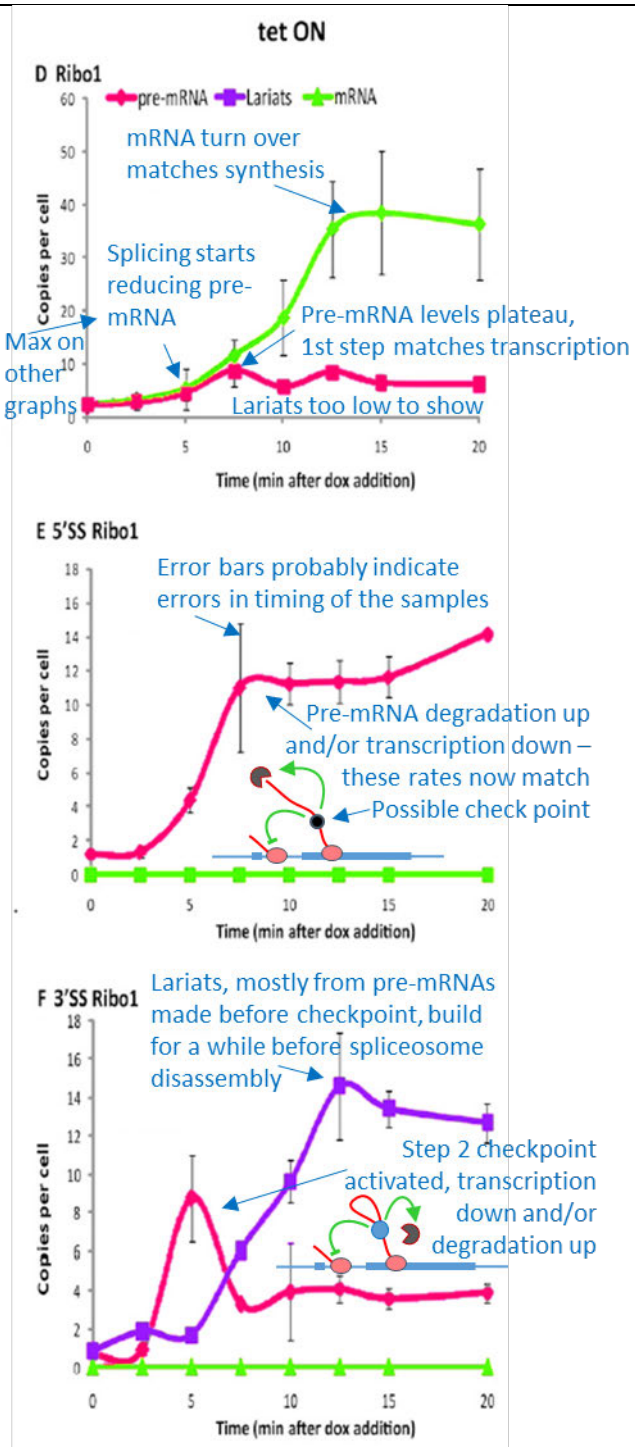


Figure 2: As P.Figure 5
Plateaus indicate that a steady state has been reached, where rates of synthesis (transcription or splicing) match those of removal (degradation or splicing).
Sub-figure numbering taken from P.Figure 5
D) Ribo1: Splicing and transcription are in balance at the 7.5 minute time point. For mRNA splicing = degradation at 12.5 minutes
E) 5'SS Ribo1: pre-mRNA accumulation is limited by 7.5 minutes, by degradation and/or reduction of transcription. Presumably, intron retention is detected, by the binding of factors like Msl5p or at translation via NMD.
F) 3'SS Ribo1: first step of splicing occurs but proceeds no further. The pre-mRNA levels decline after 5 minutes. Either an initial train of polymerases has passed or a checkpoint stimulates degradation/reduced transcription. Continuing 1st step splicing of existing pre-mRNAs, and those few newly synthesised, continue to raise the lariat levels until the stalled spliceosomes can be released allowing the lariat to be degraded.

Where the 3'SS is mutated (F), the first step of splicing can occur, but not the second. Initially pre-mRNA rises with transcription at a similar rate to Ribo1 and 5'SS Ribo1 but fails to reach the 7.5 minute peak (seen in D & E), peaking at the 5 minute time point as first step forms the lariat and removes the 5'SS.

This allows 2.5 minutes for spliceosome assembly and the first step of splicing to take place, in this, the first round of splicing. Pre-mRNA peaks but lariat continues to rise as stocks of pre-mRNA are spliced and a few more made. By 12.5 minutes the rate of transcription and 1st step splicing is matched by rates of spliceosome disassembly and lariat degradation. The peak of pre-mRNA at the 5 minute time point is intriguing, the pre-mRNA does not peak until 7.5 minutes in the other two reporters. Pre-mRNA is destroyed by the 1st step which can proceed in this reporter. The implication is that there is feedback from a second step checkpoint to reverse the accumulation of pre-mRNA by slowing transcription or increasing degradation.

2.3.4 *In situ* Reporter mRNA levels

Fluorescent *in situ* hybridisation (FISH) probes to the exon were used to give a copy number estimation independent of the RT-qPCR values. At the population level, the results with the two methods were remarkably consistent (P.Table 1A). It was also remarkable that the FISH analysis showed that induction of the reporter was not at all consistent between individual cells (P. Table 1B). Despite being an artificial induction system over 27 % of the cells either did not respond or had fewer than 5 copies of the reporter RNA, whereas some had almost 100 copies (Figure 3).

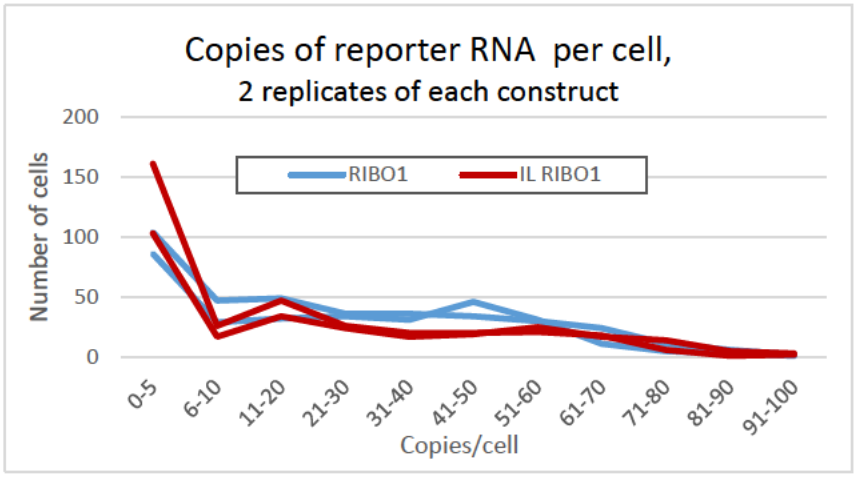


Figure 3: Distribution of copies per cell of the Ribo1 reporter RNAs
From the data in TABLE 1B

Diversity of copy number was matched by diversity of the number of cells with active transcription sites (P.Figure 3E), measured by exonic signal to sites of transcription. These sites of transcription were identified by FISH with probes to the intron. Over 20 % of cells had no detectable transcription sites. This variation is consistent with promoter cycling between on and off states where transcription factors regularly disassociate and re-associate with their binding sites.

In 73 % of cells there was only one location with intronic signal, processing was generally performed at a single site, most likely at the site of transcription and not transported or allowed time to diffuse elsewhere.

2.3.5 3‘ End Formation and De-adenylation

The 3’ end cleavage and polyadenylation sites were mapped in 51 cDNAs produced from the Ribo tetOFF strain by ligating an oligo at the 3' end of the mRNA. 85 % of these had sites 1407-1409 from the initial AUG, the rest 1396-1400. The same procedure used to generate the cDNA for cloning was used with nested PCR to determine poly(A) tail length (P.Figure 6A), both in steady state and once expression repressed (P.Figure 6B and C). Initially, at steady state the poly(A) tail was 20 to 70 adenosines (A)s long, but declined at 0.8 A/minute (1.5 A/minute for the intronless reporter). The apparent increased stability of the spliced transcript could be pre-mRNAs continuing to be spliced and entering the mRNA pool freshly polyadenylated

2.3.6 Repression and RNA Degradation

The Ribo1 tetOFF reporter was used to estimate the rates of degradation (P.Figure 4). The 5’SS and 3’SS Ribo1 RNAs both had very short half-lives (λ); 2.5 minute for 5’SS Ribo1 and 4.46 for 3’SS Ribo1.

The 3'SS Ribo1 pre-mRNA (as measured by the 3’SS RT-qPCR assay) has a λ of 4.46 minutes and the lariat at 4.32. The numbers for these two RNAs are very similar, well within the error, and could be explained if the rate limiting step is disassembly of the stalled spliceosomes; once this happens degradation of the aberrant RNA rapidly follows.

The 5'SS Ribo1 transcript does not have a spliceosome to be removed, only Msl1p and Mud2p are associated, so revealing the RNA for degradation is more rapid and the λ shorter.

The half-lives for the two viable transcripts, Ribo1 and IL Ribo, is much greater, at 8.32 and 6.20 minutes respectively. The shorter half-life of IL is unexpected, as there should be nothing distinguishing these two mRNAs once spliced. As has been suggested above, continued splicing could add new mRNAs to the pool after termination of transcription, artificially increasing half-life. In other eukaryotes the increased stability of the spliced transcript would be ascribed to the influence of the Exon Junction Complex (EJC) (Woodward et al., 2017), but this does not exist in *S. cerevisiae*. There is a recent report of the spliceosomal component, Prp17p, being found in the ribosomal mRNA fraction (particularly monosomes), (Wen et al., 2017). This protein might fulfil a similar role to the EJC and indeed its deletion abolishes splicing dependant NTC. If this is an equivalent to the EJC it could promote its RNA's stability. However, degradation of these RNAs is faster than would be predicted from the poly(A) shortening data, indicating that there are other mechanisms of controlling degradation at play.

A curious feature of the degradation of the Ribo1 and IL Ribo is that there is no degradation until after the first sample is taken, 5 minutes post doxycycline addition, whereas the two mutant RNAs degrade rapidly after addition of doxycycline. A possible explanation is that mutant RNAs are degraded co-transcriptionally (or peri-transcriptionally), whereas the viable RNA degrades post-transcriptionally. Therefore, transcripts started immediately prior to doxycycline addition would be continued until completed, some minutes later.

2.3.7 Splicing before Cleavage

If spliced transcript is detected from cDNA reverse transcribed from a primer downstream of the 3' end cleavage site, it must have been spliced co-transcriptionally. This is the basis of the cleavage assay, Figure 4A and P.Figure 7A. The very well-defined Ribo1 3' cleavage site (section 2.3.5),

allowed co-transcriptional splicing to be confidently assessed. cDNA was also reverse transcribed from an oligo dT primer to give the splicing status of RNA released from transcription by 3' end processing. The RNA species measured after reverse transcription from the cleavage primer (i.e. still transcribing RNAs) are reproduced on one graph to ease comparison (Figure 4B).

The pre-mRNA signal with the cleavage primer is basal until the 9 minute mark (P.Figure 7B), implying that this is when transcription reaches the pre-cleavage primer binding site. Comparing this with P.Figure 5D where mRNA and pre-mRNA reach similar levels at 7.5 - 8 minutes indicates that it has taken about 1 to 1.5 minutes to transcribe the approximately 1000 bp between the two RT-primer binding sites. The data for the oligo dT primer is similar but shifted one sample later (30 seconds), implying this is the approximate time taken to cleave and polyadenylate sufficient length of poly(A) tail to be bound by the oligo dT RT-primer.

The peak of pre-mRNA at 9 minutes (Figure 4B), corresponds to two samples and has returned to near basal level by the time point only 60 seconds later. The peak also rises to 8 copies in the 30 seconds since the previous time point. This is in contrast to the gentler rise of pre-mRNA seen in Figure 2, although the poorer temporal resolution of this portion of the study makes comparison difficult. This indicates that there is an initial pulse of transcription, possibly involving several polymerases. This initial pulse is not fully spliced, although a number have undergone the 1st step of splicing. This is reinforced by the polyadenylated data (P.Figure 7B), which also shows a peak of unspliced polyadenylated RNA. Any subsequent transcripts, even those arriving at the 3' end only 60 seconds later, have undergone at least the 1st step of splicing and most have completed splicing, as can be seen by the lariat-exon2 species and mRNA species rising after this point.

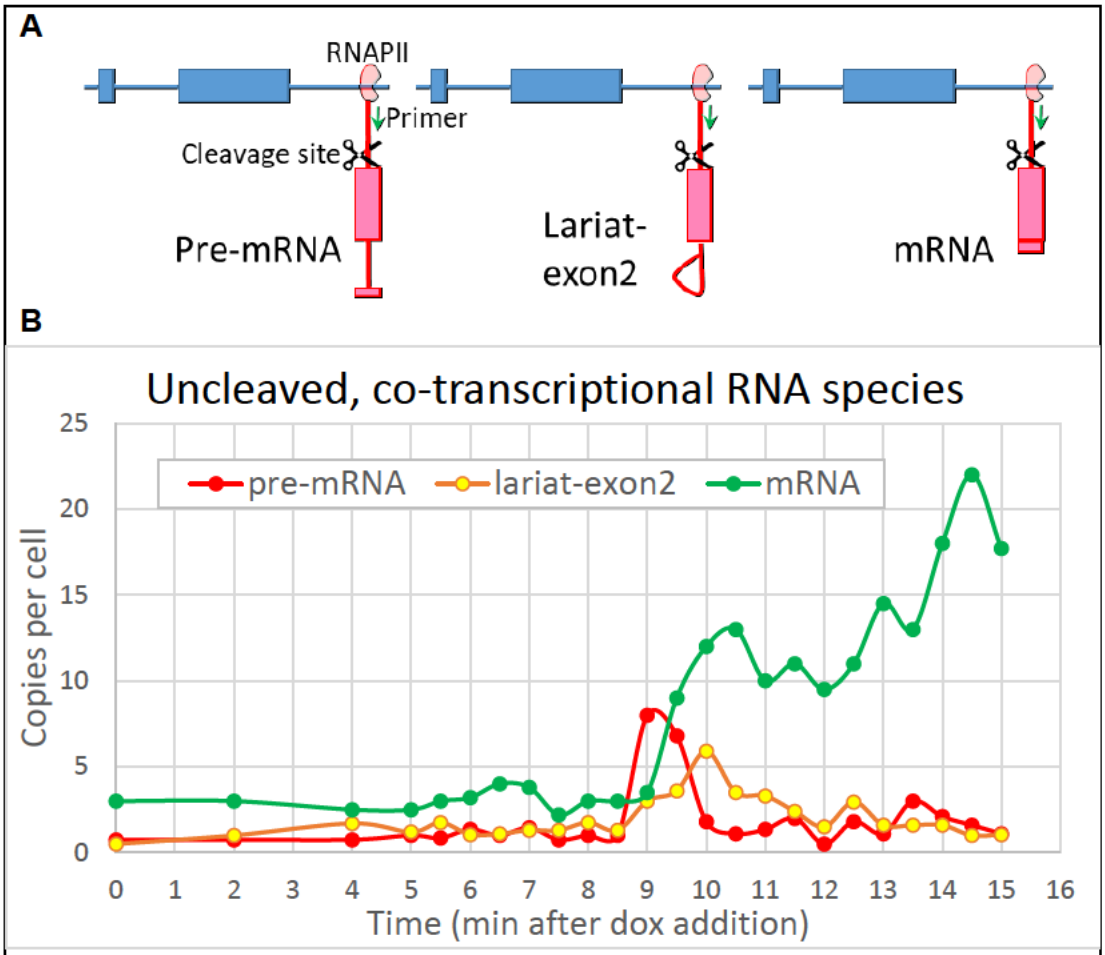


Figure 4: Co-transcriptional RNA species.
A) Any pre-mRNA or mRNA reverse transcribed from downstream of the cleavage site must be still being transcribed, therefore any processing performed on it must have been co-transcriptional.
B) Between 8.5 and 9 minutes after doxycycline addition the polymerase reaches the region where the cleavage primer binds, the cleavage primer can now bind the nascent RNA, facilitating reverse transcription and allowing the pre-mRNA to be detected. Initial transcripts, at the 9 minute time point, are not fully spliced (pre-mRNA), but some have undergone 1st step of splicing (lariat-exon2). Fully spliced mRNAs rapidly accumulate after that point. Levels of RNAs that have not been spliced at all while the polymerase is beyond the cleavage site decline rapidly, as the rate of the 1st step of splicing catches up with transcription. Second step catches up slower, as seen by the later accumulation and broader peak of lariat-exon2 between 9 and 12 minutes, before 2nd step can splice all RNAs still associated with the polymerase downstream of the cleavage site, i.e. still transcribing RNAs.
The higher levels of mRNA seen at early time points are the background of reporter production in these experiments.
The 3'SS PCR detects pre-mRNA and lariat-exon2 species so copy number for the lariat-exon2 were calculated by subtracting the pre-mRNA levels from 3'SS levels.
Reproduced from P.Figure7, see that figure for error bars, removed here to reduce clutter.

The lariat-exon2 peak occurs two time points later providing an estimation for the time to perform the second step of splicing as being in the order of 30 - 60 seconds; once first step has taken place. The broader peak of the 3'SS shows that lariat-exon2 RNA species continue to be made but the rate of the second step does not achieve parity with the 1st step for 2-3 minutes. This could indicate that second step factors also take some time to recruit and the step to reach full efficiency. Again the polyadenylated RNA fraction shows increasing splicing 30 seconds behind un-cleaved RNA as partially spliced RNAs are polyadenylated and post-transcriptional splicing occurs (P.Figure 7C).

A possible explanation for the data in Figure 4B is that the initial pulse of transcription becomes entrained by the need to remove histones and possibly to recruit elongation, splicing and 3' end formation factors. Initially this process is incomplete, and spicing is not fully co-transcriptional. Within a very short time splicing factors have been recruited and reporter mRNA splicing becomes increasingly co-transcriptional.

The cleavage assay, along with the data from FISH that intron signal is most likely lost at the site of transcription and the RT-qPCR assay showing rapid production of mRNA, are all entirely consistent with co-transcriptional splicing. This is particularly evident after 9 minutes post induction. Presumably, splicing factors must migrate to the site of transcription. It is not clear how the splicing factors arrive at the elongating pre-mRNA. It could simply be diffusion from nearby (but spliceosomal movement is slow, see introduction), active transport or even the DNA moving to a nuclear domain where splicing is more prevalent (Jerković et al., 2019). This topic would make an interesting further study.

2.4 Reinterpretation

Since the paper was written there have been new developments and reinterpretations of the results, some of which have already been mentioned. The poorer performance of the tetOFF system is probably due to residual doxycycline and metabolic changes because of centrifugation. There is evidence in the RT-qPCR data of the first wave of transcription setting up the

spliceosome (Aitken et al., 2010). The degradation rates are interesting on re-examination; evidence of co-transcriptional degradation and the likelihood that spliceosome disassembly is the limiting factor. The concept of checkpoints in splicing has also given new ways of approaching the problem of mis-splicing.

2.5 Impact

This paper describes the development of fundamental tools, and the initial results were the basis of a subsequent publication (Alexander et al., 2010). This later paper has had a major influence on the field of co-transcriptional splicing and is highly cited. Data derived from this chapter's study was also used to model transcription and co-transcriptional processing (Aitken et al., 2010).

The cleavage assay provided clear evidence for co-transcriptional splicing of the reporter transcripts, and was the forerunner, and possibly the inspiration for, nascent RNA-seq, NET-seq and SMIT.

2.6 Contribution

Work towards the Quantification of Transcripts in Populations of Cells section was developed and performed entirely by me, including the FACs analysis. I also developed the sampling strategy of placing the samples directly into alcohol on dry ice to rapidly fix cell metabolism. The PCR system for assaying splicing was developed and validated entirely by me, and I produced the RT-qPCR data for Table 1A. Some of the RT-qPCR assays used for P.Figures 4 and 5 were also performed by me, in collaboration with Dr Ross Alexander. I developed the 3’ end cleavage assay, although the data shown were produced by Dr Alexander. The RiboSys consortium met throughout the project to discuss results, interpretation and problems. I was an active member of those meetings.

All additional data presented in this chapter are my own, as is any re-interpretation of the results and conclusions from the publication.

2.7 References

Most references are in the publication, a few additional ones are listed here.

Aitken, S., Robert, M.-C., Alexander, R.D., Goryanin, I., Bertrand, E., and Beggs, J.D. (2010). Processivity and Coupling in Messenger RNA Transcription. PLoS ONE *5 (1)* e8845.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell *40*, 582–593.

Jerković, I., Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Higher-Order Chromosomal Structures Mediate Genome Function. J. Mol. Biol *432*, 676-681.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. J. Biotechnol. *75*, 291–295.

Wen, J., Marzi, L., Wang, J., Ye, J., and Brogna, S. (2017). Splicing-dependent NMD requires Prp17 in Saccharomyces cerevisiae. BioRxiv 149245.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip. Rev. RNA *8*.

2.8 Reprint

METHOD

RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*

ROSS D. ALEXANDER,[1,2] J. DAVID BARRASS,[1,6] BEATRIZ DICHTL,[3,6] MARTIN KOS,[1,6,7] TOMASZ OBTULOWICZ,[4,6] MARIE-CECILE ROBERT,[5,6] MICHAL KOPER,[4] IWONA KARKUSIEWICZ,[4] LUISA MARICONTI,[3] DAVID TOLLERVEY,[1,2] BERNHARD DICHTL,[3] JOANNA KUFEL,[4] EDOUARD BERTRAND,[5] and JEAN D. BEGGS[1,2]

[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, United Kingdom
[2]Edinburgh Centre for Systems Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JD, United Kingdom
[3]Institute of Molecular Biology, University of Zürich, CH 8057 Zürich, Switzerland
[4]Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Pawinskiego 5a, 02-106 Warsaw, Poland
[5]Institut de Génétique Moléculaire de Montpellier, CNRS UMR5535, Institut Fédératif de Recherche 3, 1919 route de Mende, 34293 Montpellier Cedex 5, France

ABSTRACT

We describe methods for obtaining a quantitative description of RNA processing at high resolution in budding yeast. As a model gene expression system, we constructed tetON (for induction studies) and tetOFF (for repression, derepression, and RNA degradation studies) yeast strains with a series of reporter genes integrated in the genome under the control of a tetO7 promoter. Reverse transcription and quantitative real-time-PCR (RT-qPCR) methods were adapted to allow the determination of mRNA abundance as the average number of copies per cell in a population. Fluorescence in situ hybridization (FISH) measurements of transcript numbers in individual cells validated the RT-qPCR approach for the average copy-number determination despite the broad distribution of transcript levels within a population of cells. In addition, RT-qPCR was used to distinguish the products of the different steps in splicing of the reporter transcripts, and methods were developed to map and quantify 3′-end cleavage and polyadenylation. This system permits pre-mRNA production, splicing, 3′-end maturation and degradation to be quantitatively monitored with unprecedented kinetic detail, suitable for mathematical modeling. Using this approach, we demonstrate that reporter transcripts are spliced prior to their 3′-end cleavage and polyadenylation, that is, cotranscriptionally.

Keywords: single-molecule FISH; RNA quantification; splicing; transcription; yeast

INTRODUCTION

In eukaryotes, most nuclear pre-mRNAs undergo processing at their 5′ ends (capping), 3′ ends (cleavage and polyadenylation), and internally (splicing to remove noncoding introns), and there is good evidence to suggest that at least some of these processes are functionally coupled to transcription (for review, see Perales and Bentley 2009). To improve our understanding of these complex events and to facilitate mathematical modeling, methods are required for the quantitative measurement of RNA processing events with good time resolution during induction or repression of transcription.

The available techniques for quantifying transcripts have inherent advantages and disadvantages. Northern blotting can provide information on the size as well as the expression level of a transcript, but measures the relative abundance only of individual mRNAs and has a limited dynamic

[6]These authors contributed equally to this work.
[7]**Present address:** Biochemie-Zentrum der Universität Heidelberg (BZH), Im Neuenheimer Feld 328, 69120 Heidelberg, Germany.
Reprint requests to: Jean D. Beggs, Wellcome Trust Centre for Cell Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK; e-mail: jbeggs@ed.ac.uk; fax: 44 131 6508650.
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2162610.

 RNA (2010), 16:2570–2580. Published by Cold Spring Harbor Laboratory Press.

range. Microarray analyses can compare transcripts genome-wide, but have a poor dynamic range and are impractical for high-resolution kinetic studies. Measurements of mRNA transcripts as copies per cell have relied heavily on hybridization-based complexity assays (Iyer and Struhl 1996), serial analysis of gene expression (SAGE) (Velculescu et al. 1997), kRT-PCR (reverse transcriptase initiated PCR-kinetic PCR) (Holland 2002), and genomic run-on (Garcia-Martinez et al. 2004). However, using competitive PCR between genomic DNA and cDNA, Miura et al. (2008) concluded that the overall abundance of the yeast mRNA population is at least twice as high as previously estimated. Another approach is to visualize and count individual transcripts in single cells using fluorescence in situ hybridization (FISH) (Femino et al. 1998; Raj et al. 2006; Zenklusen et al. 2008). This approach also found that expression levels for particular yeast genes were higher than previously estimated and can vary substantially among cells in a population (Zenklusen et al. 2008).

Nutritionally modulated promoters such as *MET3* or *GAL1* (for review, see Maya et al. 2008) are not ideal for regulated gene expression studies as the changes in growth medium are likely to result in pleiotropic effects on cell metabolism. In particular, both a galactose to glucose shift and the presence or absence of methionine were reported to impact on RNA turnover in yeast (Dichtl et al. 1997; Bousquet-Antonelli et al. 2000). An alternative is to use promoters that respond to nonmetabolizable drugs. For example, tetracycline (tet) and its analog doxycyline (dox) have no significant effects on yeast cell growth, morphology or global transcription levels (Wishart et al. 2005).

As model gene expression systems we constructed tetON (for induction studies) and tetOFF (for repression, derepression, and RNA degradation studies) yeast strains, each of which expresses tetracycline-responsive repressor and *trans*-activator proteins (dual activator/repressor regulation [Belli et al. 1998b]). Using reporter genes in these strains we developed an RT-qPCR approach to determine mRNA abundance as average numbers of copies per cell in a population, and validated this approach by comparison with single-molecule FISH that measures transcript copy numbers in individual cells. We describe PCR-based methods to monitor RNA production, splicing, 3′-end maturation and degradation in a quantitative, tightly controlled manner and with rapid response times, as is required for mathematical modeling of these events. This combination of approaches allows a much more detailed kinetic analysis of the expression of an individual gene than is generally achieved by other methods. This is illustrated by high-resolution analysis of both steps of pre-mRNA splicing and 3′-end maturation of a reporter transcript, which permits cotranscriptional splicing to be demonstrated directly. Although developed and optimized for the tet-regulated expression system in yeast, these methods should have wide utility for gene expression studies with other regulated promoters and in other organisms.

RESULTS AND DISCUSSION

Construction of strains and tetracycline/doxycycline-regulated reporter genes

Tetracycline repressor (tetR) and *trans*-activator (tTA) fusion proteins bind at promoters containing tetO DNA sequences only in the absence of tetracycline or doxycycline, and are displaced by drug addition. Conversely, the modified tetR′ and tTA′ proteins bind tetO DNA only in the presence of the drug (Gossen et al. 1995; Gari et al. 1997; Belli et al. 1998a,b). For tight control and a better dynamic range of regulated transcription, dual activator/repressor regulation was used, in which each yeast strain coproduces both tetR and tTA′ for induction (tetON) or tetR′ and tTA for repression (tetOFF) (Belli et al. 1998b; for details, see Supplemental Material 1). In the original dual expression system of Belli et al. (1998b), the regulatable tetO promoter and the tTA or tTA′ transactivator gene were present together on a plasmid and the tetR/R′-repressor gene was integrated in the yeast genome. To place reporter genes or endogenous yeast genes under tet control, it would be preferable to integrate the tet0 promoter/tTA expression cassette in the genome, in order to maintain the chromatin environment and avoid possible plasmid stability/segregation problems.

To introduce the tetracycline dual activator/repressor system into the yeast genome, the tTA or tTA′ hybrid *trans*-activator gene was integrated at the *LYS2* locus and tetR′ or tetR hybrid repressor (Belli et al. 1998b) was integrated at either the *LEU2* or *URA3* locus in three different yeast strains (for details, see Materials and Methods). For initial tests of the system the *tet07-CYC1* hybrid promoter (see Supplemental Material 2 for the sequence) was inserted in front of the nonessential *STE3* gene (Supplemental Fig. S1). In the case of the tetOFF system, strains YMK118 (derived from CEN-PK2-1C) and YMK120 (W303) gave better levels of expression than YMK119 (BY4741), and YMK120 and its *ura3* derivative, YIK120, were chosen for further study (for genotypes, see Supplemental Material 1).

In contrast, none of the corresponding tetON strains showed induction even after 15 h in the presence of dox (data not shown). After testing many constructs, finally, a functional tetON strain, YIK91, was produced with tTA′ on a centromeric plasmid expressed under the control of the strong *S. cerevisiae ADH1* promoter (P_{ADH1}tTA′), and with the tetR-SSN6 repressor gene expressed from the *Schizosaccharomyces pombe adh1* promoter (P_{adh1}tetR-SSN6 from pCM247) (Belli et al. 1998b) integrated at the *LEU2* locus of W303. The main differences between this and the Belli et al. tetON system are that the $tetO_7$-regulated gene is integrated into the genome rather than being on a plasmid, and tTA′, although still plasmid based, is expressed from the stronger *ADH1* promoter. In comparison, expressing tTA′ from the CMV promoter (Belli et al. 1998b) produced little or no expression in our strain, and the *TEF2* promoter

(Mumberg et al. 1995) supported only low level expression (Supplemental Fig. S2).

As model genes to analyze transcription, splicing, degradation, and 3′-end formation of pre-mRNA, a series of "RiboSys" reporter genes was constructed and each was integrated in the genomic *his3* locus of the tetON and tetOFF strains, under control of the $tetO_7$/CYC1-UAS promoter (Supplemental Material 2). The reporter genes are based on a previously described (Hilleren and Parker 2003) series of hybrid *ACT1/PGK1* sequences in which the *ACT1* intron was modified by inserting two copies of the λ boxB sequence (57 base pairs [bp] each), enabling it to be readily distinguished by RT-qPCR from the endogenous *ACT1* intron without affecting splicing (data not shown). Variants of the "wild-type" reporter, Ribo1, contain a point mutation at the 5′ splice site (5′SS) or 3′ splice site (3′SS) or lack the intron (IL) (Fig. 1; Hilleren and Parker 2003).

Quantification of transcripts averaged over a population of cells

Previous applications of RT-qPCR to quantify transcripts have depended on the use of reference genes for normalization; however, this is unreliable, as few genes have constant levels of expression under different growth conditions (Teste et al. 2009). In order to obtain an estimate of the absolute number of copies of a transcript per cell, several stages in the isolation and assay of the RNA need to be quantified, including cell lysis, extraction of RNA from lysed cells, reverse transcription, and qPCR efficiency (summarized in Fig. 2; for full details, see Supplemental Material 3). Of these, the first is possibly the most difficult to control. An estimate of the efficiency of cell lysis can be obtained by determining the amount of genomic DNA recovered compared with the expected amount. The difference between the observed and

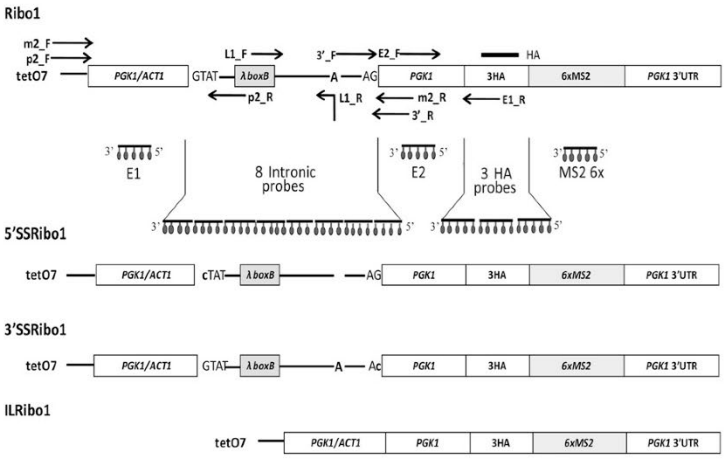


FIGURE 1. Diagram showing important features of the RiboSys reporter genes and the positions of hybridization probes and primers used in reverse transcription (RT) and RT-qPCR reactions. The reporter genes are modified from those developed by Hilleren and Parker (2003). Briefly, exon 1 contains the *ACT1* 5′ UTR and 30 codons from *PGK1* and *ACT1* sequences, and is followed by the *ACT1* intron, then exon 2, which contains ~100 bp of *PGK1* fused in frame with three copies of the haemaglutinin epitope (3HA), followed by six MS2 coat protein binding sites (6MS2) at the start of the *PGK1* 3′ UTR. To further discriminate between the reporter transcripts and endogenous *ACT1* and *PGK1* transcripts in RT-qPCR assays we inserted two copies of the lambda N (box B; 57 bp) sequence 51 bp downstream from the *ACT1* 5′ splice site. Transcription of the RiboSys reporters is under control of the $tetO_7$/*CYC1-UAS* promoter and the tetracycline (or doxycycline) regulatable dual activator/repressor system (Belli et al. 1998b). The intron is represented by a thin line with letters indicating splice signals (uppercase) and positions of splice site mutations (lowercase). Positions of flourophore probes used in FISH experiments are indicated *below* the reporter: 11 probes for mRNA and eight probes for intron (pre-mRNA) detection. Primers used for RT-qPCR assays are indicated by arrows and a northern probe by a black line. Primer pairs, p2_F and p2_R, amplify the 5′SS region of unspliced pre-mRNA. Exonic primer pairs, m2_F and m2_R, are used to quantify the spliced mRNA under conditions that do not produce a product from unspliced pre-mRNA. E2_F and E1_R measure all forms of this transcript regardless of their progress through the splicing reactions. The intron-containing products of the first and second steps of splicing are branched/lariat structures, containing a 2′-5′ phosphodiester bond that blocks the progress of reverse transcriptase from the 3′ end of the intron. The lariat intron-exon2 species can be measured in two ways. First, the product produced by primers 3′_F and 3′_R that flank the 3′SS, represents the sum of the unspliced pre-mRNA and the lariat intron-exon2 species. The difference between this and the RT-qPCR product of unspliced transcripts that crosses the 5′SS is due to the lariat intron-exon2 species. Second, oligonucleotide L1_R with the 5′ end complementary to the 5′ end of the intron and the 3′ end complementary to the intron sequence immediately 5′ of the branch site, but with a mismatch (A instead of T) opposite the branch site A (Vogel et al. 1997), anneals at the branch site of lariats and can be used to prime the RT reaction. and, with L1-F, primes RT-qPCR amplification of the circular part of the lariat. This measures the sum of the lariat-exon product of the first splicing reaction and the excised intron lariat product of the second splicing reaction. In normal conditions, the level of lariat species is very low, as the lariat intron-exon2 product of the first splicing reaction is rapidly processed in the second step and the excised intron is rapidly debranched and degraded. However, if the second splicing reaction is slow or blocked or if there is a defect in release and degradation of the excised intron, these species can be quantified. The sequences of all oligos are listed in Supplemental Table S3.

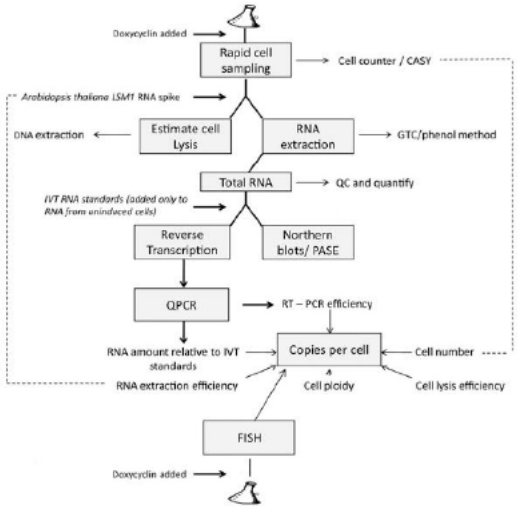


FIGURE 2. Scheme of work for estimating RNA copy number by determining the efficiency of cell lysis, RNA extraction, RT-qPCR, and FISH imaging. Briefly, the cell number is determined and the efficiency of cell lysis is estimated based on the amount of DNA recovered (total DNA is assayed spectroscopically and individual genes by RT-qPCR) compared with the expected amount of DNA, assuming that in an exponentially growing population of haploid cells the average cell contains 1.4 genomes worth of nuclear DNA. The *S. cerevisiae* genome is ~0.012 pg (Fungal Genome Size Database www.zbi.ee/fungal-genomesize). Total RNA purified from the yeast lysates is measured optically. The efficiency of RNA recovery is estimated by measuring the recovery of a known amount of nonyeast RNA (in our case, *Arabidopsis thaliana LSM1*) added to the lysate. The reverse transcription and qPCR stages are standardized by comparison with known amounts of in vitro transcribed RNAs with the same sequence, mixed with total RNA purified from cells that do not express the reporter genes. The results from this procedure are compared with numbers obtained by visual inspection of individual transcripts detected by FISH in large numbers of single cells. The details of the various procedures are given in the Supplemental Material.

the theoretical yield of DNA is largely due to lack of cell lysis and so an estimate of the efficiency of cell lysis can be derived.

An estimate of the efficiency of RNA recovery from the cell lysate was obtained by adding a known amount of nonyeast in vitro transcribed RNA, and determining by RT-qPCR how much was recovered at the end of the RNA purification. This could only provide an estimate as it was not certain that the exogenous RNA (with no proteins bound to it) will behave in the same manner as endogenous mRNA from newly lysed yeast cells. The reverse transcription and qPCR reactions were standardized by comparison with known amounts of in vitro transcribed RNAs with the same sequences, after mixing with total RNA extracted from cells that lack the reporter gene.

The positions of hybridization primers used in the reverse transcription (RT) and qPCR reactions are shown in Figure 1, together with an explanation of how the products of the different stages of the splicing reaction were distinguished. This included an assay for branched, lariat RNA species that accumulate if the second splicing reaction is slow or blocked, or if there is a defect in release and degradation of the excised intron. It was shown previously by microarray analysis that doxycycline (at 10 times the concentration used here) has no significant effect on global transcription levels in yeast (Wishart et al. 2005). Therefore, assaying an endogenous transcript that is unaffected by dox provides a measure of the reproducibility of this technique during the time course of an experiment (Supplemental Fig. S3).

Image-based measurement of mRNAs in individual yeast cells

In order to determine transcript abundance in single cells a FISH approach was employed, using Cy3-labeled deoxyoligonucleotide probes complementary to the reporter transcripts. Previous studies have shown that single RNA molecules can be efficiently detected as isolated spots in microscopic images, using five to 10 oligo probes (Femino et al. 1998). Two sets of probes were used. The first set detects the intron, and consists of a pool of eight probes that cover the entire length of the Rib1 intron. The second detects the exons, and contains a pool of six probes. One probe hybridizes six times against the MS2 tag, and a maximum of 11 oligonucleotides from this pool can thus hybridize at the same time on the mRNA or pre-mRNA. The exonic pool of probes labeled small and bright spots in the nucleus and the cytoplasm, which correspond to isolated, individual mRNA molecules (Femino et al. 1998). A semiautomated procedure was then developed to detect and automatically count single molecules of the reporter transcripts (seen as bright individual spots) in 3D stacks of cells expressing the various reporters and nonexpressing control cells (Fig. 3). Briefly, a Gaussian filter of 0.300 μm was applied to the images to remove variation in intensities smaller than the size of single molecules. All local maxima were then determined and a spot was defined as a sphere of the desired diameter (0.3 μm in this case), centered on each maximum. To remove the background, the spots were then thresholded on the basis of their local contrast (the standard deviation of the signal in the volume of the spot). For this, we used a control strain that did not contain the reporter gene, and selected a minimal local contrast that left less than one spot per cell in the control strain. Eye examination of the results indicated that this procedure automatically detected single RNA molecules with a reasonable efficiency. To quantify this better, we plotted histograms of the number of spots detected as a function of local spot contrast (see Fig. 3B). With the entire exonic probe set, the spots detected in Rib1 expressing cells were of much higher local contrast than in nonexpressing cells. Using a contrast threshold that excluded most spots in the nonexpressing cells (Fig. 3B, yellow vertical bar) (the threshold is chosen such that one molecule per cell is detected in the control), the specific spots could be sorted from the background. In addition, by overlaying normalized histograms for the control and Rib1 expressing strains, it could be seen that

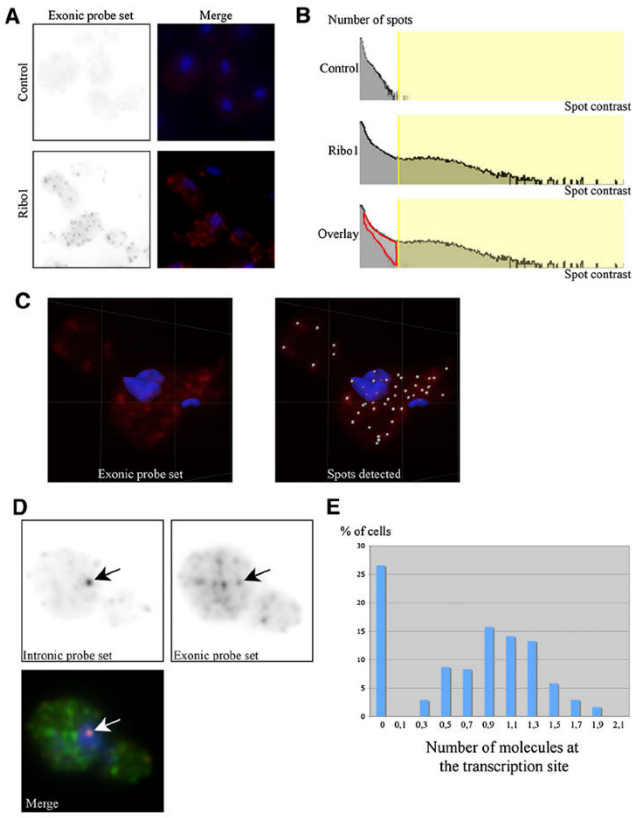


FIGURE 3. Image based measurement of Ribo1 RNAs in individual yeast cells. (*A*) Detection of single molecules of the Ribo1 reporter RNA in TetOFF strains. Control strain lacking the reporter gene (*top*), or expressing Ribo1 (*bottom*) were hybridized in situ with the exonic probe set. *Right* panels display maximal image projections of the RNA signal (red) overlayed with the nuclei (blue). Each field is a projection of a 3D stack (6 × 6 × 6 μm). (*B*) Efficiency of RNA detection. Histograms of the number of spots versus spot contrast across an entire 3D stack (63 × 63 × 6 μm) are shown for control and Ribo1 expressing cells. The area shaded in yellow corresponds to the spots included in the analysis after thresholding. *Bottom* panel: overlay of the two histograms revealing the amount of Ribo1 mRNA molecules lost by the thresholding procedure (red area; <20% of the total number of spots identified) (*C*) Maximal image projection (3.5 × 3.5 × 6 μm) of a cell with the spots identified indicated with small spheres on the *right* panel. (*D*) Single plane of a Ribo1 expressing cell with both the exonic (Cy5, green) and the intronic signals (Cy3, red). Blue: Dapi. Each field is 3.6 × 3.6 μm. Arrows indicate the position of the intronic focus that identifies the putative transcription site. (*E*) Histogram plotting the number of molecules of Ribo1 RNA, identified with the exonic probe set, at the intronic foci (putative transcription site). Cells with zero molecules have no signal for the intronic pool of probe.

the threshold excluded a small fraction of spots from the Ribo1 sample (Fig. 3B, outlined in red). By measuring the total number of specific spots (the ones above the threshold and those outlined in red), we could estimate that <20% of the specific spots were lost by the thresholding procedure, yielding a detection efficiency higher than 80%. It should be noted that this approach is limited to RNAs with a copy number above 1, which is similar to the threshold chosen here, and that above 60 per cell many fluorescent spots merge, making identification of individual RNA molecules less reliable.

In order to compare the two methods of estimating copy numbers for Ribo1 and ILRibo1, duplicate cultures of the tetOFF strains were propagated in one lab, where the RT-qPCR method was used, and samples of these cultures were fixed with paraformaldehyde and sent to be assayed by FISH in another lab. The numbers obtained by RT-qPCR and by FISH are in good agreement (Table 1A), with the difference being similar to the variation seen when comparing duplicate cultures using a single method. Of course the RT-qPCR approach measures the copy number averaged over the entire population of cells, and the FISH approach shows that the copy number varies greatly from cell to cell, with the distribution being from 0 to 100 copies per cell for Ribo1 and ILRibo1 (Table 1B). This is likely due to the promoter cycling between On and Off states, as previously proposed and modeled (Blake et al. 2006; Zenklusen et al. 2008; Aitken et al. 2010). It therefore seems likely that much of the variability in copy-number measurements between duplicate cultures is due to biological variation. Furthermore, cultures grown in rich (YPD) medium showed a much greater variation between biological replicates than cultures grown in synthetic medium (data not shown). This suggests that the frequency of promoter cycling between On and Off states may be sensitive to the growth rate or metabolic state of the cells, which may be induced by subtle differences between batches of rich media.

Consistently, FISH with the intronic probes showed that 73% of these haploid cells had a single focus of signal, which was always in the nucleus, and which colocalized with a spot labeled by the exonic probes and was thus most likely the transcription site (Fig. 3D). To quantify the number of RNA molecules present at the intronic foci, we first calculated the average intensity of the cytoplasmic spots labeled by the exonic probe set, which corresponded to single, full-length mRNA molecules. We then divided the intensity of the spot that colocalized with the intronic foci by this value. This showed that the signal corresponded to one molecule of fully transcribed pre-mRNA (Fig. 3E), consistent with RT-qPCR

TABLE 1A. Copy-number measurement by RT-qPCR and FISH using the same cultures

	tetOFF strains (no dox)	
Reporter	RT-qPCR culture A/B	FISH culture A/B
Ribo1 mRNA	21.6/19.4	25.8/26.3
ILRibo1	24.4/22.9	26.0/18.0

quantification of 1.28 ± 0.1 copies of pre-mRNA per cell in the tetOFF strain when derepressed, and indicating that only a few polymerases were loaded on the gene at any time. Moreover, the lack of intronic signal elsewhere in the cells was compatible with splicing taking place at the transcription site, cotranscriptionally.

To determine whether unspliceable pre-mRNAs are retained at a particular area in the cell, we analyzed the localization of the 5′SSRibo1 mutant by in situ hybridization. Using the intronic pool of probes, we found that the mutant unspliced pre-mRNAs accumulated at the nuclear periphery (Supplemental Fig. S4), as previously described (Galy et al. 2004).

Repression and RNA degradation kinetics

A particular advantage of tetOFF strains is the ability to measure RNA degradation rates following the turn off of expression, by simply adding dox without the need to change the growth medium. By northern analysis, the Ribo1 and ILRibo1 mRNAs had similar apparent half-lives of 9.4 and 8.8 min, respectively (Fig. 4A), compared to 8.3 and 6.2 min when assayed by RT-qPCR (Fig. 4B) (measured as the slope from the maximum at 5 min). The RT-qPCR assay seems likely to be more reliable, as more replicate assays were performed and the assay was standardized against known amounts of synthetic RNA standards. The 5′SS and 3′SS mutant reporter transcripts are defective in the first and second step of splicing, respectively, and have much shorter apparent half-lives (Fig. 4C,D). Interestingly, degradation begins immediately after dox addition, whereas there is a delay of several minutes before degradation of Ribo1 and ILRibo1 mRNAs is apparent. This most likely reflects very rapid targeting of the aberrant, unspliced transcripts for degradation by the RNA surveillance systems (Bousquet-Antonelli et al. 2000; Hilleren and Parker 2003; Houseley and Tollervey 2009).

Knowledge of the number of molecules per cell and RNA half-life permit calculation of the transcript synthesis rate, at least for unspliced RNAs (Perez-Ortin et al. 2007). For ILRibo1 this gives values of 1.9–2.5 and 4.9–6.3 (molecules/cell/min), respectively, at steady state in the tetOFF and tetON strains. Assuming that initiation of transcription is the limiting step in the rate of synthesis, this indicates an initiation event every 27 and 11 sec, respectively (Iyer and Struhl 1996), reflecting the generally higher copy numbers in the tetON strains.

Derepression and induction kinetics

The validated RT-qPCR method for copy-number measurement was then used to follow the kinetics of derepression and induction in the tetOFF and tetON strains, respectively. Note that the RT-qPCR assay of Ribo1 exon2 in the tetOFF strain (Fig. 5A) (exon) reaches slightly higher levels than that of Ribo1 mRNA, and approximately equates to the amount of mRNA plus pre-mRNA and lariat-exon2 (i.e., all the RNA species that contain exon2), supporting the specificity and reliability of the RT-qPCR assays of the different RNA species. In the tetOFF strain under repressed conditions (Fig. 5A) (0 min) less than one copy per cell of Ribo1 mRNA could be detected. Higher levels of reporter transcripts were reproducibly detected in the tetON strains (Fig. 5D–F), in both uninduced (0 min) and induced conditions, and with more rapid accumulation than in the tetOFF strains (Fig. 5A–C; note the different scales). The unspliced 5′SSRibo1 transcripts (Fig. 5B,E) reached lower maximum copy numbers than for Ribo1 mRNA, which may be explained by the faster turnover of this species (Fig. 4C). With 3′SSRibo1 expression (Fig. 5C,F), the lariat-exon 2 product of step 1 splicing was readily measured, but a pronounced transient accumulation of pre-mRNA suggests that there is a delay before the first step of splicing occurs, apparently more so than with Ribo1.

With both tetON and tetOFF strains the lag period between dox addition or removal, respectively, and first detection of transcripts varied slightly between cultures grown at different times. The reason for this is not known but, combined with the variation in the copy number noted above, may indicate sensitivity of the transcription machinery to

TABLE 1B. Distribution of copies per cell

	Cell number			
	Ribo1		ILRibo1	
Number of copies	A	B	A	B
0–5	104	86	103	161
6–10	47	29	17	26
11–20	49	32	34	47
21–30	36	34	24	26
31–40	36	31	17	20
41–50	34	46	19	20
51–60	30	31	25	21
61–70	24	11	17	18
71–80	11	5	14	6
81–90	6	3	5	1
91–100	2	1	3	2

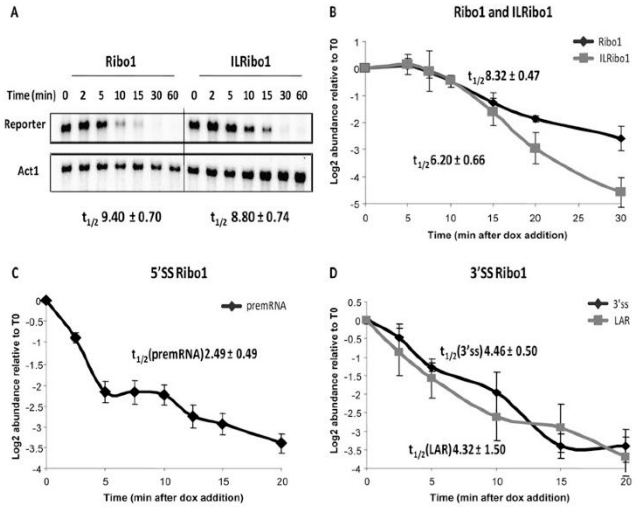


FIGURE 4. Kinetics of repression and turnover of Ribo1 reporter transcripts using the RiboSys tetOFF strains. The tetOFF strains with intron-containing Ribo1 or intonless ILRibo1 (*A*,*B*), mutant 5′SSRibo1 (*C*), or mutant 3′SSRibo1 (*D*) reporters were grown in SD minimal medium with doxycycline (4 μg/ mL) added at time 0 to initiate repression of transcription. (*A*) Northern blot analysis of Ribo1 and ILRibo1 reporters showing relative mRNA levels following repression. Apparent half-lives of the reporter transcripts are shown below the blot. (*B–D*) Repression of RiboSys reporters showing relative abundance ($\log_2$) of the Ribo1 and ILRibo1 mRNAs (*B*), 5′SSRibo1 pre-mRNA (*C*), and 3′SSRibo1 lariat-exon2 intermediate (*D*) assayed by RT-qPCR of the 3′SS region (3′SS) or using lariat-specific RT-qPCR (LAR). The apparent half-lives, calculated from the linear parts of the curves, are displayed.

subtle differences in growth medium, other environmental conditions or the growth phase of the cells.

3′-End formation

In yeast, the 3′ ends of all mRNAs are produced by cleavage and polyadenylation. The efficient formation of the mRNA 3′ ends is important for mRNA stability as well as for capping, splicing, transcription, and translation (Minvielle-Sebastian and Keller 1999). Polyadenylation of the Ribo1 transcripts is driven by 3′ UTR sequences derived from the *PGK1* gene. Initially, using total RNA obtained from tetOFF strains grown under nonrepressive conditions, we determined the sites of poly(A) addition, using a ligation mediated cDNA amplification strategy to determine sequences encompassing the poly(A) sites (Fig. 6A). Analysis of 51 independent clones revealed a major poly(A) site located at positions 1407–1409 (relative to the *PGK1* start codon) that was used in 85% of cases and a minor site at positions 1396–1400 was used in 14% of the sequenced clones. Next, we used the same assay to follow the poly(A) status of reporter transcripts during derepression in tetOFF strains (Fig. 6B). The poly(A) length ranged from approximately 20 to 70 adenosines (A's) at steady state (30 min and later). The abundance of transcripts carrying longer or shorter tails, respectively, dropped sharply beyond those values. Linear poly(A) shortening during the early phase of reporter derepression was used to deduce the apparent rate of deadenylation, which was found to be 0.8 adenosines min^{-1} for the Ribo1 reporter and 1.5 A's min^{-1} for ILRibo1 (Fig. 6C). These values are consistent with the deadenylation rate of 2–3 A's min^{-1} that was reported for the *PGK1* 3′ UTR in the context of the related PGK1pG construct (Tucker et al. 2001). However, the mRNA half-lives associated with the Ribo1 reporters are short (<10 min) (see Fig. 4) compared to *PGK1* mRNA that is very stable ($t_{1/2}$ 45 min) (Parker and Jacobson 1990). The lack of correlation between deadenylation rates and mRNA half-lives indicates that deadenylation-independent mechanisms may determine the turnover of Ribo1 reporter transcripts.

High-resolution kinetic analyses reveal splicing prior to 3′-end cleavage

There has been much debate about the extent to which splicing takes place co- or post-transcriptionally in yeast (Gornemann et al. 2005; Lacadie and Rosbash 2005; Tardiff et al. 2006). However, the assays used to date, including chromatin immunoprecipitation to measure co-transcriptional recruitment of splicing factors or of RNA-binding proteins that distinguish spliced from unspliced transcripts, are indirect and therefore may be inaccurate. The modified RT-qPCR approach described here permits more quantitative measurements of transcript accumulation and both steps of splicing. As RNA species are measured in copies per cell, different RNA species can be compared with each other and between different samples and cell cultures. Furthermore, identification of the positions of 3′-end cleavage/polyadenylation permits RT-qPCR analysis of the kinetics of 3′-end formation and how this relates to splicing. As shown in Figure 7, by taking samples at short time intervals (30 sec), it is possible to detect a peak of full-length nascent (pre-mRNA) transcripts at 9 min, prior to their 3′-end cleavage, which takes place within 30 sec. The lariat-exon2 product of the first step of splicing can also be detected at 9 min, prior to 3′-end processing. Furthermore, at 9.5 min spliced mRNA that is not yet 3′-end processed can be detected, prior to the appearance of polyadenylated mRNA at 10 min. These data clearly demonstrate that splicing takes place on a large proportion of transcripts prior to 3′-end cleavage, as was suggested by the FISH analysis; and we

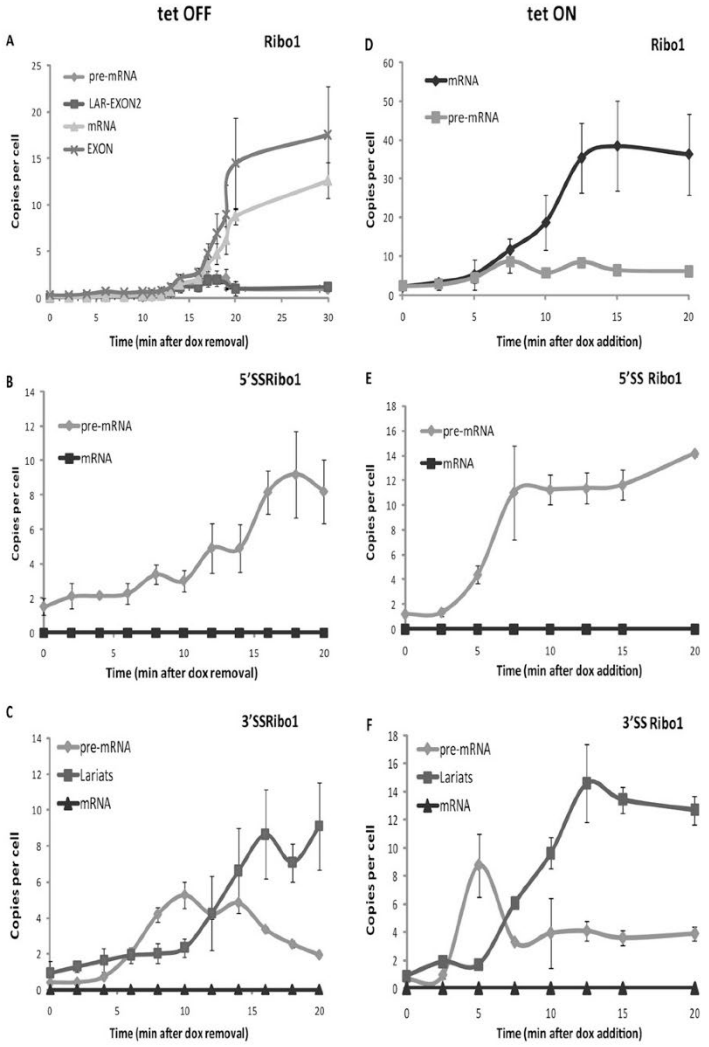


FIGURE 5. Kinetics of transcription and splicing of RiboSys reporter transcripts in tetOFF and tetON strains. A time course is shown of the levels of pre-mRNA (pre), lariat intron-exon2 splicing intermediate (lar-exon2; measured by 3′SS assay and subtracting the amount of unspliced pre-mRNA) (see Fig.1), exon2 (exon) and spliced mRNA in copies per cell. (*A–C*) tetOFF cultures were grown in SD medium to midlog phase (OD600 0.5) with transcription repressed by the presence of dox (4 μg/ mL). After 90 min of transcriptional repression, cells were harvested by centrifugation and resuspended into medium without dox (time 0). (*D–F*) tetON cultures (YIK91 with various Ribo1 reporters integrated at the *his3* locus) were grown in SD minimal medium (-trp) with doxycline (4 μg/ mL) present to induce transcription. (*A,D*) Ribo1; (*B,E*) 5′SSRibo1; and (*C,F*) 3′SSRibo1. In each panel the data represent three experiments, each assayed in triplicate, and error bars indicate standard deviation.

conclude that Ribo1 transcripts are, to a large extent, spliced cotranscriptionally.

Furthermore, the half-life of each step of splicing can be approximately estimated as 30 sec, equating to 60 sec for the conversion of pre-mRNA to mRNA. This is similar to a measurement based on indirect ChIP analyses in yeast (Tardiff et al. 2006) and significantly faster than the 5–10 min measured for splicing an intron in human cells, irrespective of intron length (Singh and Padgett 2009). Tardiff et al. (2006) determined that splicing begins when transcription is >600 bp beyond the 3′SS and that maximal cotranscriptional splicing requires ~1 kilobase (kb) past the 3′SS. The Ribo1 major and most distal 3′ cleavage site was mapped at 735–737 nucleotides (nt) past the 3′SS. It can also be seen that cleavage and polyadenylation take place within 30 sec of transcription through this region. This compares with an estimated 55 sec for 3′-end cleavage and polyadenylation of HIV-1 (Human Immunodeficiency Virus type 1) mRNA, measured by FRAP (fluorescence recovery after photobleaching) (Boireau et al. 2007). This level of quantification and kinetic resolution will permit mathematical modeling to determine more accurate rates for the two steps of splicing and estimates of how much splicing occurs pre- and post-transcriptionally.

CONCLUSIONS

The regulated genes studied here are chromosomally located, rather than being on plasmids as in previous tet-regulated constructs (Belli et al. 1998b). Yen et al. (2003) developed a similar tetOFF strain with $tet0_2$ or $tet0_7$ promoters integrated in the genome but, to our knowledge, this is the first report of such a tetON strain. The tet-regulated expression of other genes could be studied by integrating other ORFs with the $tet0_7$ promoter at the *HIS3* locus, using the pMK121 or pMK123 plasmids that we describe. The $tet0_7$ promoter could also be integrated upstream of any gene in the genome, although the efficiency of tet-repression is variable with different yeast genes (Mnaimneh et al. 2004; Wishart et al. 2006), and in our experience this is also true for tet-regulated induction.

The RT-qPCR approach for estimating RNA copy number could also be used to measure the regulated expression of endogenous genes from their native promoters and, in principle, it should be adaptable to other organisms, including the

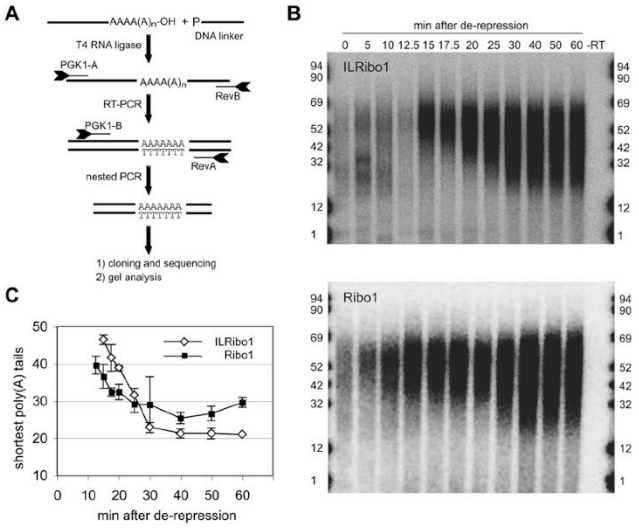


FIGURE 6. 3′-End formation and deadenylation (*A*) Flow diagram of the PCR-based method that was applied to analyze transcript 3′-end formation. In a first step a 5′ phosphorylated DNA linker oligonucleotide is ligated by T4 RNA ligase to available 3′-OH groups. This is followed by first-strand cDNA synthesis using the linker-specific RevB oligonucleotide and subsequent PCR analysis. The obtained PCR product is further amplified by nested PCR using RevA and PGK1-B oligonucleotides. The obtained PCR product can be transferred into a plasmid and analyzed by DNA sequencing. Alternatively, the nested PCR can be performed with radioactively labeled oligonucleotides and PCR products can be separated on denaturing polyacrylamide gels. (*B*) Time course of adenylation/deadenylation following derepression of ILRibo1 and Ribo1 in the tetOFF strain, analyzed as described in *A*. Radioactive PCR products were resolved on 6% (w/v) polyacrylamide/ 8.3 M urea gels. The length of the poly(A) tract associated with the PCR products is indicated by numbers. To control for specificity of PCR amplification a minus reverse transcriptase (−RT) control was included. (*C*) Graph depicting the length of the shortest poly(A) tails observed at individual time points during reporter derepression. Data are the mean of three (− Intron) and four (+ Intron) experiments and error bars represent S.D. Deadenylation rates were derived from initial linear phases of poly(A) shortening.

analysis of tissue-specific gene expression. Using this approach to estimate the copy number of constitutively produced transcripts requires careful consideration of the synthetic RNA used to standardize the RT-qPCR assays. As the efficiency of RT-qPCR differs for purified transcripts compared to transcripts assayed in cell extracts, ideally the in vitro transcribed RNAs should be distinguishable from the endogenous transcripts without affecting the properties of the qPCR reaction, although it may be possible to use a synthetic RNA that is identical to the endogenous transcript and measure the increase in abundance.

The absolute measurement of RNA levels facilitates modeling studies of RNA metabolism, for example, allowing comparisons between different growth conditions, strains, or cell types. Previous efforts to model RNA processing pathways in yeast have depended on northern hybridization, other gel-based analyses (Cao and Parker 2001, 2003) or imaging (Zenklusen et al. 2008). The use of RT-qPCR is more sensitive and offers a greater dynamic range than these approaches, and the ease of scale permits more replicates and therefore a better measure of accuracy and reproducibility. As illustrated here, the more quantitative RNA measurements combined with rapid sampling can produce very-high-resolution data for in vivo kinetic studies, revealing kinetic details of RNA processing events not previously detected in yeast. By comparing pre-mRNAs with different introns, and the effects of *trans*-acting mutations that affect different stages of RNA processing, this approach should permit the identification of rate-limiting steps in the production and turnover of intron-containing transcripts and distinguish whether the kinetics of transcription and splicing are functionally coupled.

MATERIALS AND METHODS

Yeast strains, plasmids, and growth conditions

Full details of yeast strains, plasmids, and primer sequences are described in Supplemental Material 1. Sequences of the tet-CYC1 promoter and reporter constructs are in Supplemental Material 2. Briefly, to construct the tetOFF strains, the gene encoding the tTA hybrid *trans*-activator was integrated at the *LYS2* locus, deleting *LYS2*. For tetOFF, the gene encoding tetR′ hybrid repressor was integrated along with *URA3* as a marker at the *ura3* or *leu2* locus. For the successful tetON strain YW156, the gene encoding tetR hybrid repressor, expressed from the strong *Schizosaccharomyces pombe adh1* promoter, was integrated at the *leu2* locus, and the gene encoding the tTA′ hybrid *trans*-activator was cloned with the strong *S. cerevisiae ADH1* promoter on a centromeric plasmid. Plasmids pMK121 and pMK123 were constructed that contain the $tetO_7$/*CYC1-UAS* promoter followed by a multiple cloning site, plus Nat (pMK121) or KanR (pMK123) as a yeast transformation marker, all bordered by *HIS3* flanking sequences, allowing integration at the *his3* locus.

Cultures were grown in synthetic dropout (SD) medium to midlog phase, transcription was repressed or induced by the addition of doxycycline to 4 μg/mL and aliquots were snap-frozen by pipetting into an equal volume of methanol sitting on dry ice. The frozen cells in the methanol slurry were pelleted by centrifugation and stored at −80°C.

RT-qPCR measurements

An overview of the scheme of work for measuring the efficiency of cell lysis, RNA extraction, and RT-qPCR normalization for RNA copy-number estimation is shown in Figure 2. Full details of the various procedures are given in Supplemental Material 3. Primer sequences are listed in Supplemental Material 1. Transcription

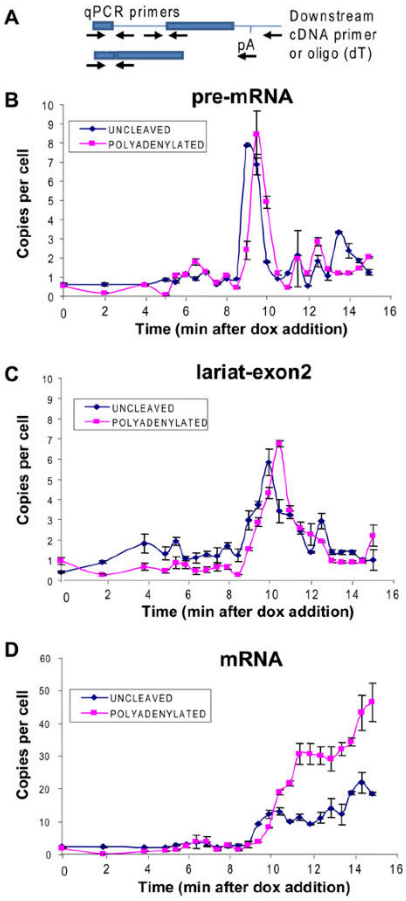


FIGURE 7. Kinetic analysis of splicing and 3′-end formation of Ribo1 transcripts demonstrates cotranscriptional splicing. RNA, sampled at 30-sec intervals during a time course of induction of Ribo1, was reverse transcribed with either oligo (dT) to copy cleaved and polyadenylated transcripts, or with a primer downstream of the mapped cleavage/poly(A) sites to copy uncleaved transcripts (there is a short delay between transcription through the poly(A) site and 3′-end cleavage). qPCR was performed to measure unspliced and spliced transcripts. (*A*) Schematic of the assay, showing the approximate positions of primers used for qPCR of unspliced pre-mRNA (across 5′SS), lariat intron-exon (across 3′SS then subtract the value for pre-mRNA) and spliced mRNA (across exon junction) or used to prime cDNA synthesis. pA indicates the 3′-end cleavage/polyadnylation site. Results are shown for: pre-mRNA (*B*), lariat intron-exon2 (*C*), and mRNA (*D*). As a control, it was demonstrated that no cDNA was produced when oligo (d'T)-selected polyadenylated RNA was reverse transcribed using the downstream primer (data not shown).

profiles obtained using calculation of RNA copies per cell were confirmed by measuring the changes in reporter level compared to the quantity of TDH1 mRNA (that is unaffected by doxycycline) and also by comparison with Northern blot analysis (Supplemental Fig. S5). This verifies that including efficiencies of cell lysis and RNA recovery in the former approach does not misreport the end result.

Image-based measurement of mRNA in individual yeast cells

Yeast strains were grown in SD medium and fixed with paraformaldehyde in midlog phase. Probes were 40–55 nt long and each contained four amino-allyl dT that were conjugated to Cy3 (Femino et al. 1998). The sequence of the probes and the protocols for in situ hybridization and to detect and automatically count single mRNA molecules are described in Supplemental Material 4.

Polyadenylation status analysis

The assay (for overview, see Fig. 6A) is described in detail in Supplemental Material 5.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

This work was funded by EC grant no. LSHG-CT-2005-518280 for the RiboSys project and the Wellcome Trust, and benefitted from facilities in the BBSRC-funded Edinburgh Centre for Systems Biology. J.D. Beggs holds the Royal Society Darwin Trust Research Professorship.

Authors' contributions: R.D. Alexander—construction of the RiboSys reporter genes, copy number, and kinetic data by RT-qPCR; Figures 1, 2, 4B–D, 5, 7 and Supplemental Figure S4. J.D. Barrass—development of the copy-number estimation and lariat RT-qPCR methods. M. Kos and D. Tollervey—construction of the tetOFF strains and extensive efforts to produce tetON strains; Supplemental Figures S1, S2. M.-C. Robert and E. Bertrand—FISH analyses; Figure 3 and Supplemental Figure S5. Beatriz Dichtl, L. Mariconti, and Bernhard Dichtl—3′-end cleavage sites and PASE assays; Figure 6. T. Obtulowicz, M. Koper, I. Karkusiewicz, and J. Kufel—construction of the successful tetON strain, analyses by northern and qPCR comparing with TDH1; Figure 4A and Supplemental Figure S6. J.D. Beggs—coordination of the project and preparation of the manuscript.

Received March 5, 2010; accepted September 8, 2010.

REFERENCES

Aitken S, Robert MC, Alexander RD, Goryanin I, Bertrand E, Beggs JD. 2010. Processivity and coupling in messenger RNA transcription. *PLoS ONE* **5:** e8845. doi: 10.1371/journal.pone.0008845.

Belli G, Gari E, Aldea M, Herrero E. 1998a. Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. *Yeast* **14:** 1127–1138.

Belli G, Gari E, Piedrafita L, Aldea M, Herrero E. 1998b. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Res* **26:** 942–947.

Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* **24:** 853–865.

Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, Pradet-Balade B, Backer V, Kornblihtt A, Marcello A, Bertrand E. 2007. The transcriptional cycle of HIV-1 in real-time and live cells. *J Cell Biol* **179:** 291–304.

Bousquet-Antonelli C, Presutti C, Tollervey D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102:** 765–775.

Cao D, Parker R. 2001. Computational modeling of eukaryotic mRNA turnover. *RNA* **7:** 1192–1212.

Cao D, Parker R. 2003. Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell* **113:** 533–545.

Dichtl B, Stevens A, Tollervey D. 1997. Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes. *EMBO J* **16:** 7184–7195.

Femino AM, Fay FS, Fogarty K, Singer RH. 1998. Visualization of single RNA transcripts in situ. *Science* **280:** 585–590.

Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbass U. 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116:** 63–73.

Garcia-Martinez J, Aranda A, Perez-Ortin JE. 2004. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* **15:** 303–313.

Gari E, Piedrafita L, Alea M, Herrero E. 1997. A set of vectors with tetracycline-regulatable promoter system for modulated gene expression in *S. cerevisiae*. *Yeast* **13:** 837–848.

Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. 2005. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* **19:** 53–63.

Gossen M, Freundlieb S, Bender G, Muller G, Hillen W, Bujard H. 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268:** 1766–1769.

Hilleren PJ, Parker R. 2003. Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. *Mol Cell* **12:** 1453–1465.

Holland MJ. 2002. Transcript abundance in yeast varies over six orders of magnitude. *J Biol Chem* **277:** 14363–14366.

Houseley J, Tollervey D. 2009. The many pathways of RNA degradation. *Cell* **136:** 763–776.

Iyer V, Struhl K. 1996. Absolute mRNA levels and transcriptional initiation rates in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **93:** 5208–5212.

Lacadie SA, Rosbash M. 2005. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. *Mol Cell* **19:** 65–75.

Maya D, Quintero MJ, de la Cruz Munoz-Centeno M, Chavez S. 2008. Systems for applied gene control in *Saccharomyces cerevisiae*. *Biotechnol Lett* **30:** 979–987.

Minvielle-Sebastian L, Keller W. 1999. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr Opin Cell Biol* **11:** 352–357.

Miura F, Kawaguchi N, Yoshida M, Uematsu C, Kito K, Sakaki Y, Ito T. 2008. Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics* **9:** 574. doi: 10.1186/1471-2164-9-574.

Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng WT, Zhang W, Yang X, Pootoolal J, Chua G, Lopez A, et al. 2004. Exploration of essential gene functions via titratable promoter alleles. *Cell* **118:** 31–44.

Mumberg D, Muller R, Funk M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156:** 119–122.

Parker R, Jacobson A. 1990. Translation and a 42-nucleotide segment within the coding region of the mRNA encoded by the *MAT α1* gene are involved in promoting rapid mRNA decay in yeast. *Proc Natl Acad Sci* **87:** 2780–2784.

Perales R, Bentley D. 2009. "Cotranscriptionality": The transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36:** 178–191.

Perez-Ortin JE, Alepuz PM, Moreno J. 2007. Genomics and gene transcription kinetics in yeast. *Trends Genet* **23:** 250–257.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4:** e309. doi: 10.1371/journal.pbio.0040309.

Singh J, Padgett RA. 2009. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128–1133.

Tardiff DF, Lacadie SA, Rosbash M. 2006. A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly post-transcriptional. *Mol Cell* **24:** 917–929.

Teste MA, Duquenne M, Francois J, Parrou JL. 2009. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Molecular Biology* **10:** 99. doi: 10.1186/1471-2199-10-99.

Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* **104:** 377–386.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW. 1997. Characterization of the yeast transcriptome. *Cell* **88:** 243–251.

Vogel J, Hess WR, Borner T. 1997. Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res* **25:** 2030–2031.

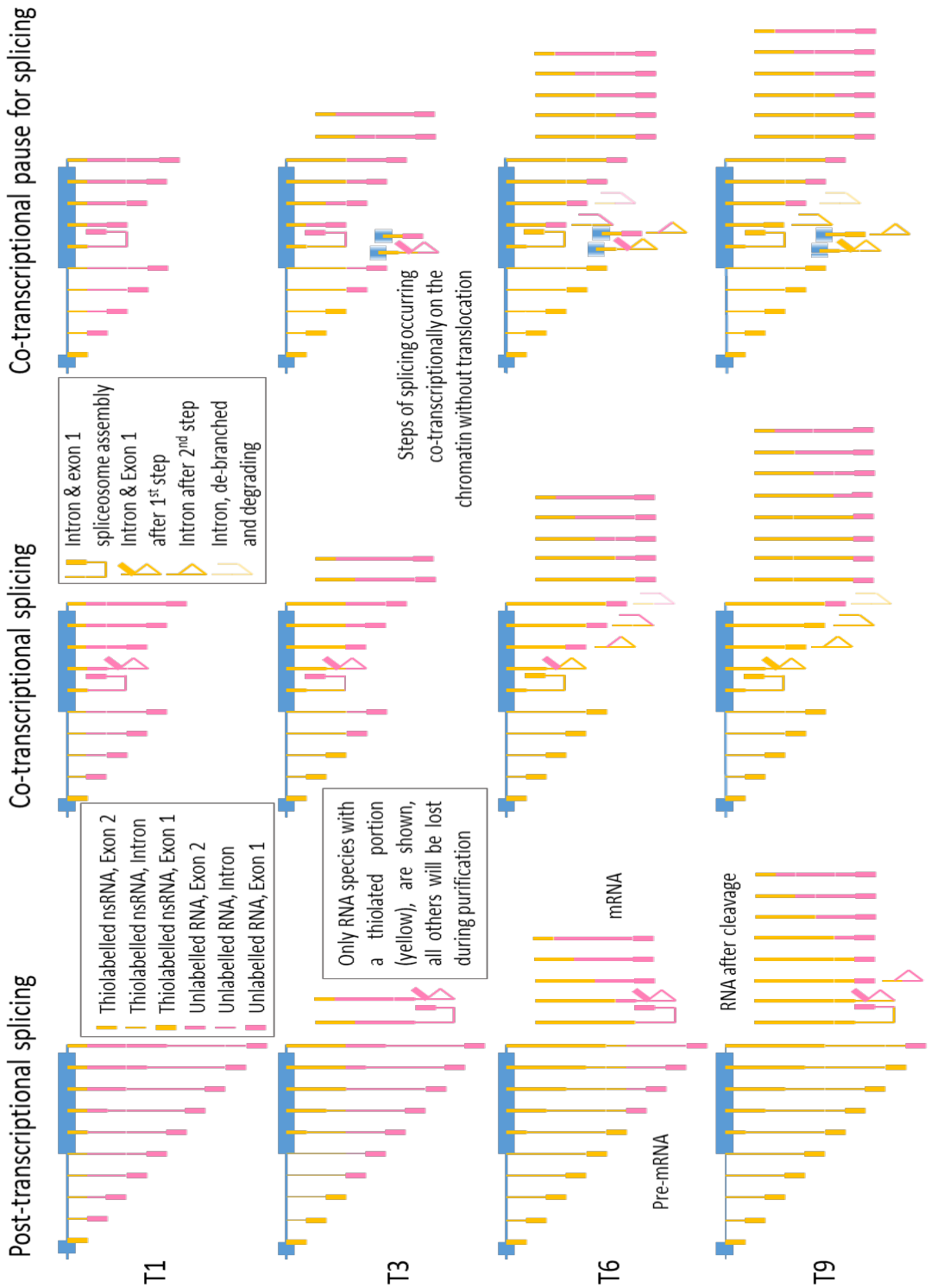
Wishart JA, Hayes A, Wardleworth L, Zhang N, Oliver SG. 2005. Doxycycline, the drug used to control the tet-regulatable promoter system, has no effect on global gene expression in *Saccharomyces cerevisiae*. *Yeast* **22:** 565–569.

Wishart JA, Osborn M, Gent ME, Yen K, Vujovic Z, Gitsham P, Zhang N, Ross MJ, Oliver SG. 2006. The relative merits of the $tetO_2$ and $tetO_7$ promoter systems for the functional analysis of heterologous genes in yeast and a compilation of essential yeast genes with $tetO_2$ promoter substitutions. *Yeast* **23:** 325–331.

Yen K, Gitsham P, Wishart J, Oliver SG, Zhang N. 2003. An improved tetO promoter replacement system for regulating the expression of yeast genes. *Yeast* **20:** 1255–1262.

Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15:** 1263–1271.

3 Metabolic Labelling of RNA Using Thiouracil



Model of thio-labelling and splicing. The blue bar is an intron-containing gene, bars coming from it are RNA, yellow portions are thio-labelled & red unlabelled. Models of post- and co-transcriptional splicing along with a pause for splicing

3.1 Research Articles

This chapter is based on the peer-reviewed publication: Barrass, J.D., Beggs, J.D., 2019. Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). J Vis Exp (Journal of Visualized Experiments).

Figures from this publication will be referred to as P.Figure followed by the number.

This is a methods paper and video describing the thio-labelling procedure in exquisite detail. Development of this procedure will be discussed along with its limitations.

3.2 Aim

The aim of metabolically thio-labelling RNA is to be able to purify, sensitively and specifically, nascent or newly synthesised RNA (nsRNA).

3.3 Experiments

The nucleotide analogue, 4-thouracil (4tU), is readily taken up by cells and incorporated into RNA as it is transcribed *in vivo*. This allows isolation of RNA produced during the brief period of labelling by covalently coupling a biotin moiety and affinity purifying using streptavidin coated beads. The more specific the purification of the newly synthesised thiolated RNA from pre-existing RNA, the less background of processed RNA one will have. Achieving a good yield of newly synthesised RNA and a low background of old RNA makes shorter labelling times possible and increases the temporal resolution.

All aspects of the protocol were optimised with the aim of improving robustness, usability, sensitivity and to reduce background. Protocols optimised for sensitivity are often themselves sensitive to slight variations in protocol, user, or environment; they are neither robust nor easy to perform. Making the protocol robust and reducing background often had the benefit of improving signal over noise which is more desirable than outright sensitivity.

3.3.1 Development of the Thio-labelling Protocol

There were several challenges to overcome in order to purify the RNA transcribed during very short labelling times. The thio-modified base must be rapidly imported into the cells and in sufficient quantities to be incorporated efficiently into the RNA. Any side effects of a nucleotide analogue on cell metabolism should be minimised. The cell metabolism, especially transcription, RNA processing and degradation must be stopped rapidly so that the duration of labelling is accurate. RNA extraction should be fast, efficient and not damaging to the RNA. The yield of purified, newly synthesized RNA should be optimal to obtain a high signal to noise ratio.

Optimisation and development of the protocol were performed using the principles of Good Manufacturing Practice.

3.3.1.1 Labelled Nucleotide Analogue

Labelling with 4tU proved more successful than with 4-thiouridine (4sU) or other thio-modified nucleobases as the source of the thio groups (Figure 1). The reasons for this are unclear, but the high solubility of 4tU in 1M NaOH reduces the toxic effect of the labelling process compared with organic solvents that are required to solubilise some nucleotide and nucleoside analogues.

The presence of 4-thiouracil in growth media at concentrations greater than 50 µM for long periods has been observed to disrupt nucleoli (Burger et al., 2013). However, the 10 µM concentration used in this protocol, and the extremely short labelling times, minimises deleterious effects while still yielding enough RNA for downstream analysis (P.Figure 1a) (Burger et al., 2013). The addition of a non-labeled (thio-free) nucleoside chase eliminated any reduction in growth rate.

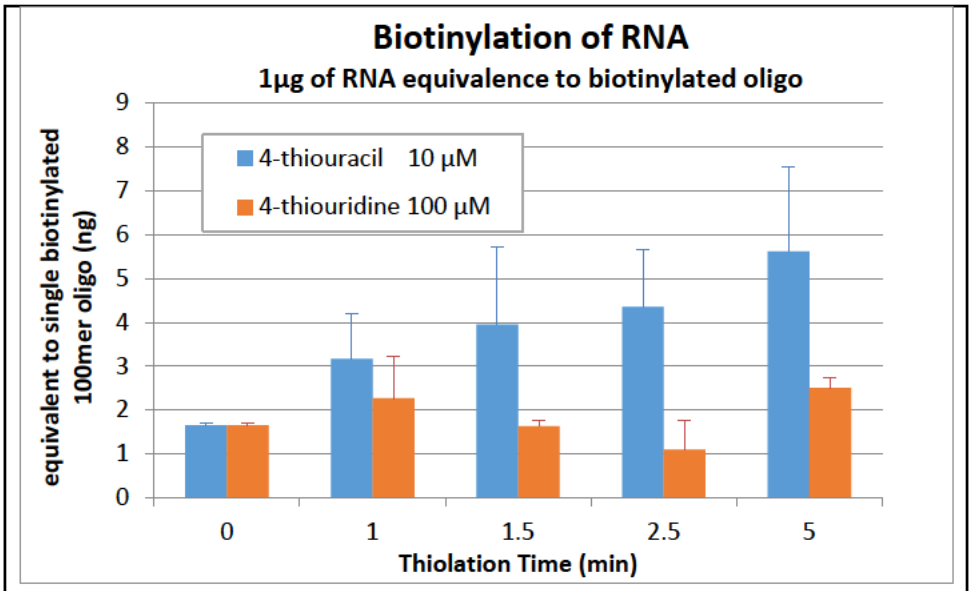


Figure 1: Comparison of thiolation with 4-thiouracil and 4-thiouridine
Total RNA biotinylated as (Barrass and Beggs, 2019) (thiouridine dissolved in H_2O). Degree of thio-labelling compared with a biotinylated oligo (Dölken et al., 2008). 4-thiouracil labels more efficiently, it is $^{1}/_{50}^{th}$ the price, less is needed and it is more stable.

In chase experiments, uridine is preferred over uracil as the chasing chemical, as it is much more soluble in water and 1 M solutions can be made. A 500-fold excess of uridine to 4tU can be added to a growing culture as a small volume of aqueous solution.

3.3.1.2 Nucleotide Importers

Although yeast cells have endogenous permeases to import uracil, such as Fui1p (Zhang et al., 2006) and Fur4p (Chevallier, 1982), for the extremely short labelling times used here, Fui1p, Fur4p or the human permease hENT (Miller et al., 2011) were over-expressed to boost uracil uptake (data not shown). Fui1p over-expressed from the endogenous promoter on a high copy number plasmid was preferred as, in addition to uracil, it also efficiently imports uridine for chase experiments.

Fui1p and Fur4p both have a PEST motif, targeting the protein for degradation once there is sufficient uracil in the cell (Marchal et al., 2000). For the chase experiments it is important that these importers available on the cell surface to

import the chase uridine. Therefore, in the plasmid-encoded gene the Fui1p PEST motif was inactivated by two point mutations.

3.3.1.3 Rapid Sampling

It is important with short thiolation times that cell metabolism is stopped rapidly, and for this methanol on dry ice was used. Methanol is preferred over ethanol as the growth medium/methanol mix has a lower freezing point, and methanol is generally cheaper. Methanol also preserves the nucleic acid content better than other methods at the cost of membrane integrity, a benefit in this application (Hobro and Smith, 2017). To speed up the sampling time as much as possible, the sample was poured directly into the fixative.

Initial experiments involved large sample sizes, often 400 ml of a log phase culture. This proved expensive and difficult to handle. Following optimisation of the RNA extraction method much smaller sample volumes could be handled, fixed more effectively and extracted much more efficiently, resulting in higher yield of nsRNA per ml of sample taken. 30 ml of sample was chosen as conveniently fitting in a 50 ml centrifuge tube containing 20 ml of methanol. This allowed much shorter sampling intervals, down to 15 seconds (or 10 with two operatives), see Figure 2.

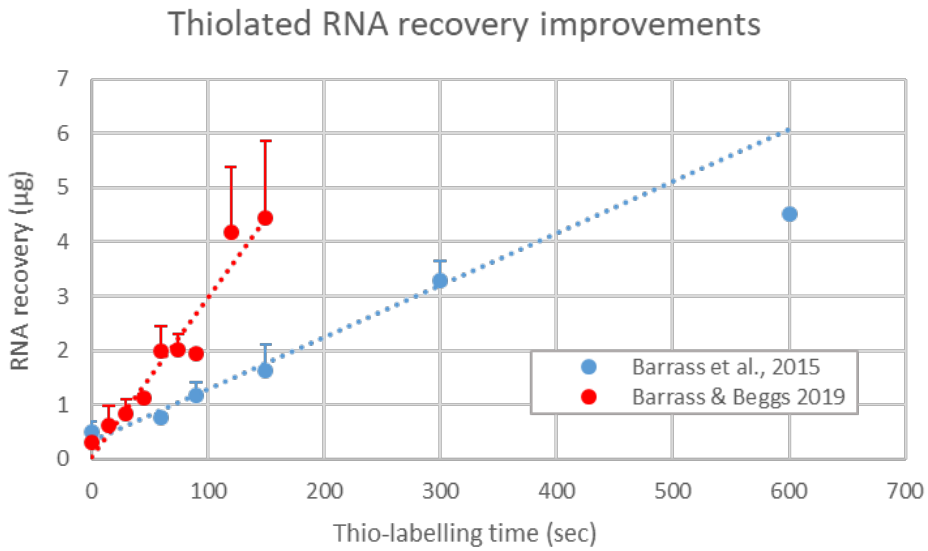


Figure 2: Improvements in the thio-labelling protocol between 2015 and 2019 Total thiolated RNA recovered from the experiments described in (Barrass and Beggs, 2019) (this chapter), and (Barrass et al., 2015) (chapter 4). Note that 400 ml of culture was used in 2015 and just 30 ml in 2019.

3.3.1.4 RNA Extraction

Extraction of total RNA can be done by many methods, although those based on chaotropic, thio-containing compounds were discounted to prevent it competing with thiolated RNA for biotin. For recovery of mRNA, LiCl precipitation is preferred, as tRNAs are less efficiently precipitated. It is advantageous to remove tRNAs as their rapid transcription and naturally thiolated nucleotides (Gustilo et al., 2008) results in much of the biotinylated RNA being tRNA. The reduced recovery of tRNA could be due to poor precipitation efficiency of small RNAs and/or those with extensive secondary structure. Therefore, if small, highly structured RNAs are of interest, traditional, alcohol based RNA precipitation methods are recommended.

3.3.1.5 nsRNA Biotinylation

For recovery of thiolated RNA, biotin is attached via the thio groups. As detailed in this protocol the use of HPDP-biotin was preferred. MTS-biotin is an alternative that can increase RNA recovery (Duffy and Simon, 2009), but is unsuitable for this protocol as it also increases the background. The biotinylation reaction was performed at 65 °C to reduce precipitation, seen with some batches of HPDP. If a permanent attachment between the RNA and biotin is desired maleamide-biotin could be used, in which case the optional TCEP reduction is strongly recommended in the protocol.

The magnetic rack seen in the video is of my own design and manufacture and superior to commercially available racks.

Whatever the biotinylation reagent, free biotin must be removed prior to further purification and this was achieved by a combination of a size exclusion column, another LiCl precipitation and multiple ethanol washes.

3.3.1.6 nsRNA Purification

The biotinylated RNA is affinity purified on streptavidin-bound magnetic beads. This protocol is similar to others reported previously (Dölken et al., 2008) but again intensively optimised to reduce background. To this end low binding centrifuge tubes are recommended, glycogen used to block the beads, and a

combination of detergents and salts in the sample buffer to minimise non-specific binding of RNA to the beads. The background of previously transcribed RNA cannot be completely eliminated due to thio-modified bases that occur naturally in some tRNAs and residual non-specific RNA binding.

The RNA is eluted from the beads by reducing the disulphide bond between the RNA and the linker to the biotin, releasing the original thiolated nsRNA. High concentrations of beta-mercaptoethanol, in excess of what is needed to release the RNA, were used to better preserve the RNA. The RNA is then precipitated and stored until required.

3.3.2 Types of Experimental Design

There are several experimental designs that can be used in the performance of a thio-labelling experiment, as explained in the paper. One approach involves continuous labelling, where 4tU is added and samples taken at intervals. This type of experiment can follow the kinetics of RNA creation and processing; an example is P.Figure 3B where splicing of a pre-mRNA is monitored.

Another experimental design uses discontinuous labelling. In this type of experiment a change in growth conditions is induced at time 0 and at intervals a portion of culture is labelled for a consistent period of time. This is illustrated best in P.Figure 2 and the animation accompanying it. Note, a growing culture is split into as many cultures as will be required, the cells are allowed to recover from the shock of being decanted before starting the labelling. These sub-cultures are identical to each other as is possible and all flasks and media pre-warmed. The results of this type of experiment are presented in P.Figure 3d, where a protein has been depleted using the β-est AID system (see chapter 6); the culture was labelled for 1 minute at different times and changes to the RNA processing monitored as levels of the target protein decreased.

Another approach is a pulse-chase experiment (P.Figure 3C), which can be used to monitor processing and decay. A short labelling with 4tU is followed by a chase with a large excess of uridine (not thiolated). A possible additional

type combines a pulse-chase with a change in growth. In this design the pulse is delivered before the change in growth and the chase along with the change. This technique is very powerful for monitoring effects that happen rapidly on the change of condition. For longer term changes, the pulse could be delivered at intervals after that change, but the practicalities of performing this experiment would make it very challenging.

3.4 Results

3.4.1 Growth in 4-thiouracil

4tU is toxic and causes slower growth (P.Figure1a) (Burger et al., 2013), the concentration used here, 10 µM, does affect growth, but a lower concentration does not lead to effective thio-labelling. P.Figure 1a shows growth over several doublings (OD_{600} 0.1 - 0.6). It is not known exactly how 4tU affects cell metabolism and growth. If the main effect is on rRNA processing, it is unlikely that splicing is affected in the labelling time scales advocated in this study. If there is a systemic effect due to, for example, weaker hydrogen bonding involving S-H groups rather than O-H groups (Mielcarek and Dołęga, 2016), then splicing could be directly affected.

3.4.2 RNA Recovery

RNA yield was remarkably consistent (P.Figure 1b, inset) and nsRNA recovery increased linearly with labelling time. Figure 2 shows that 10 minutes of labelling was saturating. The bioanalyser traces (P.Figure 1A main graph), show an absence of tRNAs (the smallest peak is 5S rRNA), due to LiCl precipitation. The rRNA peaks are visible at time point 0, so either it has not been possible to remove all total RNA contamination or some rRNA is thiolated too. rRNA and its intermediates build up rapidly. The build-up of these intermediates shows no sign of decreasing so rRNA processing could be affected by 4tU, perhaps this is work for the future. The increase of signal between the rRNA peaks is due to mRNAs, including pre-mRNA. Compared to rRNA build-up these intermediates do show signs that an equilibrium is being reached.

3.4.3 Newly Synthesised Splicing Intermediates

RT-qPCR analysis with the set of *ACT1* PCRs on the nsRNA revealed splicing of *ACT1* pre-mRNA *in vivo*, P.Figure 3b, c and d. In all cases, the level of each RNA species is normalised to the level of that RNA at steady state. As there is so much more mRNA in steady state RNA the result is further normalised to exon, to produce a plot where all species can be seen relative to one another.

The experiment in P.Figure 3b is a simple thio-labelling and thio-labelled RNA accumulates. In comparison to the steady state the pre-mRNA and lariat achieve a level where production matches removal. For both these species this is 45 seconds after the labelling begins. If we allow about 15 seconds for 4tU to be imported, the ribose, 3 phosphates added, and finally transport to the nucleus, this would indicate a time for first step of approximately 30 seconds. Lariats have very much the same profile as pre-mRNA implying that the first step (probably spliceosome assembly) is the rate-limiting step.

Plot c is the result of thio-labelling with a pulse chase. The culture was labelled for 25 seconds and then chased with a 500-fold excess of uridine. 45 seconds (i.e. 20 seconds after the end of the pulse), is required for the maximum level of pre-mRNA. After this time, splicing removes the pre-mRNA faster than transcription makes labelled pre-mRNA (if we ignore degradation). This provides an estimate of the first step of splicing taking in the order of 20 to 35 seconds. Notice that the lariat plateaus about the same time as the pre-mRNA. The pre-mRNA RT-qPCR target spans the BS, overlapping with the 3'SS PCR product, so the time taken to transcribe that small distance is probably irrelevant. The lariat peak persists slightly longer and only declines after another 30 seconds compared to the pre-mRNA. This tipping point is where production by the first step is exceeded by removal after the second step. Thus, the second step of splicing and lariat release takes approximately 30 seconds after first step. The lariat peak is broader as the level of lariats is the result of 3 processes: transcription, 1st step and de-branching, all with variable times. Both the pre-mRNA and lariat decline slowly, so there must be variability in the rates at which individual transcripts are processed. The rates of decline

of pre-mRNA and lariat are superficially similar, once more the most likely explanation is that the first step of splicing is rate-limiting.

P.Figure 3d was produced in an experiment where the second step splicing factor Prp16p was depleted and is discussed in chapter 7.

3.5 Modelling Splicing

A simulation of thio-labelling and splicing was created in MS Excel. The principle is illustrated on the front page of this chapter. Essentially the length of each segment indicates the amount of RNA recovered. Although crude it does have some power and many parameters can be changed to assess their effect on nsRNA production and recovery, Figure 3.

The output of the simulations should be compared with P.Figure 3b. Clearly the experimental data do not agree with the post-transcriptional splicing model. The two co-transcriptional models, including where transcription pauses to allow splicing to happen, are similar. The dip in the co-transcriptional lariat and pre-mRNA at the last point is due to the exon signal continuing to rise while the pre-mRNA and lariat plateau at time point 9 (T9). This is the point at which an entire virtual transcript is thiolated, the intermediate species are at their maximum labelling but the mRNA is a stable species and continues to raise mRNA and exon levels. If a longer transcript or mRNA turnover was modelled this dip would be reduced. In the pausing model exon build up is slower as the RNA is not being thio-labelled while paused. The model assumes that splicing occurs immediately after the 3'SS is transcribed, if this is not the case and splicing can occur at any time after the 3'SS has been transcribed (albeit close), then the peak at T6 and drop at T9 will be flattened out making the simulation match experimental data more closely. Splicing is certainly co-transcriptional for *ACT1*.

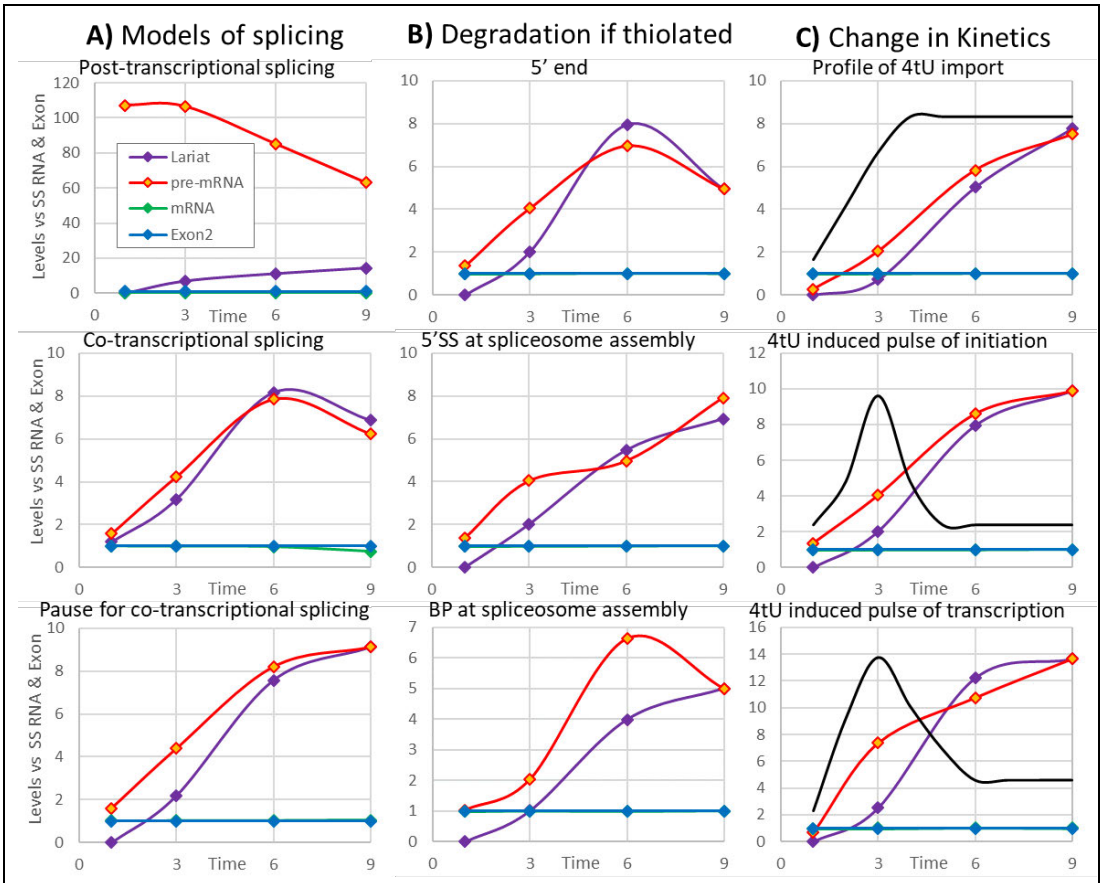


Figure 3: Thiolation and Splicing Model

Modelling precursors, intermediates and products of splicing. All data were normalised to steady state (SS RNA) and exon levels. Relative abundance of regions equivalent to the RT-qPCR products are plotted. All levels and times arbitrary units

A) Simulations of the 3 models of splicing: post-transcriptional, co-transcriptional and co-transcriptional with a pause. The pause is modelled to occur once the 3'SS has been transcribed and to allow both steps of splicing to occur

B) Models of RNA degradation related to the presence of thio nucleotides at the locations and stages indicated. These are based on the pause for co-transcriptional splicing model, although very similar traces are observed for the co-transcriptional splicing model.

C) Factors that affect kinetics, modelled on the pause for co-transcriptional splicing model. These factors are indicated by the black line on the plot (arbitrary units). Top: delayed thio-UTP delivery to the site of transcription. Middle: pulse of productive transcription initiation on the addition of 4tU (depressing initiation also had no effect). Bottom: Pulse of elongation on addition of 4tU, polymerases starved of uridine could increase elongation if uracil is added. Decreasing expression had little effect.

The pulse-chase, P.Figure 3c, would suggest there is a degree of flexibility in timing of splicing, but a more subtle model is required to test this. In the

pausing model the lariats at T1 are not detectable. This is because any lariat containing species is paused and so cannot be thio-labelled. At later time points RNA that has been labelled is spliced so lariats can be detected. A pause for both steps of splicing is modelled, a pause for just first step or spliceosome assembly would yield label lariats at T1 and would therefore fit the data better. By comparing the simulations and data it is not possible to tell if there is a pause or not, but one to allow events after 1st step is unlikely.

Figure 3B models a situation where thiouracil at various sites (5' end of the transcript, 5'SS and BS) on the transcript leads to decay of that transcript. None of the modifications tested, of which examples are shown here, either affected the profiles or matched what is seen in the data. Thio-labelling causing degradation of RNA is not something that needs to be considered when analysing the data. A similar conclusion can be drawn on stabilising the thio-labelled RNA (not shown).

However, when the time taken for the 4tU to be imported, ribose added, phosphorylated and transported to the site of transcription is taken into account, Figure 3C top, the simulations match the data slightly more accurately, so this is a factor that needs to be considered.

A pulse of transcription initiation could arise with the addition of uracil into a culture lacking uracil. However, this does not affect the simulation, here any increase in initiation is cancelled out by the exon normalisation, which also experiences an increase in initiation. This is unlikely to be an important factor (Figure 3C bottom). Neither does assuming a decrease in transcription rate match the experimental data (not shown).

One parameter for which this model has been beneficial is assessing the contamination of the purified sample by pre-existing RNA. Despite efforts to minimise this there is still significant carry-over of unlabelled mRNA. An estimation of the degree of contamination is provided in chapter 4. Given the massive abundance of total RNA, only a tiny portion of this needs to be bound non-specifically or even naturally thiolated to affect the results for mRNAs in

the nsRNAs fraction. The levels of splicing intermediates, and so all transient RNA species enriched in thiolated RNA, are not significantly affected by contamination as their levels are so low in steady state. This insight must be accounted for in future work.

From the model; splicing is co-transcriptional, and transcription may pause, but that it is unlikely to be after step 1. Levels of exon and mRNA need to be treated with caution as there is contamination from unlabelled RNA. However, comparative studies can still be performed. Factors that affect the 5' end, such as degradation or initiation rate change are controlled for by exon and steady state normalisation. Thio-related degradation of mRNA does not seem to occur, and neither does a change of transcription rate due to an influx of uracil. From this it can be tentatively concluded that pre-mRNA metabolism is not affected by the presence of 4tU and the slower growth is most likely due to problems in rRNA processing, possibly pseudouridylation.

3.6 Re-interpretation

3.6.1 Comparison with Other Methods

For direct kinetic measurements of RNA transcription and processing, various approaches have been used to distinguish newly transcribed RNAs from pre-existing species.

3.6.1.1 Reporters

One approach is to use inducible reporter genes with low background of uninduced transcription, monitoring the status of the nascent RNA by RT-qPCR or fluorescent probes that distinguish RNA intermediates, see chapter 2. Kinetic information can be obtained from immediately after induction until steady state is reached, which is essentially when transcription is in equilibrium with degradation of the transcript, this may take only a few minutes. However, the results are limited to that reporter gene (or to closely related sequences), that can be constructed and then introduced into cells. The reporter must be distinguishable from endogenous yeast sequences, at least a modified yeast

gene. It also loses much of its utility once steady state has been reached, when it becomes no more informative than an endogenous gene.

3.6.1.2 Run-On

For genome-wide studies, nuclear run-on transcription, or Genomic (or Global) Run-On (GRO) (García-Martínez et al., 2004) detects transcripts in the process of being elongated. This is often performed with purified nuclei, with sarkosyl added to pause RNAPII and inhibit new initiation events. The sarkosyl is removed and the polymerase allowed to run-on, producing transcripts from previously engaged polymerases. The new transcripts are labelled with ^{32}P-labelled nucleotides or modified nucleotides, such as 5-bromouridine triphosphate (BrUTP) (Core et al., 2008) and detected and affinity-purified with anti-BrUTP antibodies. In a variation of this approach, transcription elongation is stalled by addition of 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) prior to addition of the label, then the DRB is removed by washing, allowing transcription to resume. A significant disadvantage of these approaches is that the chemical treatments may disrupt regulation of transcription and/or downstream processing events and, more generally, disturb cell physiology.

3.6.1.3 Other Metabolic Labelling Technologies

Metabolic labelling of RNA has a long history of using radionuclide reporters. For example, inorganic ^{32}P was fed to mammals and the RNA extracted and analysed (Hokin and Hokin, 1954). A major drawback of radionuclide studies is that the labelled RNA cannot be extracted from the unlabelled, so analysis of the RNA is limited to obvious parameters such as amount and changes in length.

Click-iT (Life technologies) is an alternative, commercially available, metabolic labelling technique that can label RNA with an ethylene-modified uracil, added to the culture medium. This has the advantage that ethylene-modified nucleotides do not occur naturally, so background is low. Biotin can be covalently bound to the ethylene group but the linkage to the biotin is permanent, precluding recovery of the RNA for analysis of quality, and reducing the efficiency of reverse transcription. Also, the reagents are much

more expensive than for the 4tU protocol described here. Finally, the presence of the bulky ethylene-modification on nucleotides could have effects on the cell's metabolism.

Other studies have used thio-modified uridine or uracil, as reviewed by Tani and Akimitsu, (2012). Some have claimed “ultra-short” labelling for 5 minutes, appropriate for the kinetics of RNA processing to be measured in human cells, (Windhager et al., 2012), or labelling 2.5 minutes or more to measure transcription and degradation rates in budding yeast (Neymotin et al., 2014). However, none of these approached the 30 second resolution needed to analyse fast RNA processing events such as splicing in budding yeast.

To measure mRNA decay rates in budding yeast, Munchel et al. (2011) followed a prolonged thio-labelling with a chase that involved spinning cells down and resuspending them in new growth medium, a perturbation that has the potential to subtly alter gene expression (see Figure 4). The protocol in this chapter involves a 25 second pulse with low amounts of 4tU that can easily be overwhelmed in chase by the addition of a small volume of un-thiolated uridine without having to perform a time consuming and perturbing exchange of growth medium.

A more recent technique modifies the thiolated nucleotide so that it is detected as cytosine in high throughput sequencing (Riml et al., 2017; Schofield et al., 2018). This recoding reaction potentially allows analysis of nsRNA without purification. However, recoding requires all RNA to be sequenced, whereas purified nsRNA can be assayed by quicker and cheaper techniques, such as RT-qPCR. Another factor to consider is that, as all the RNA is sequenced, low abundance RNA species labelled for a short time will give rise to very few sequence reads, unless they can be enriched first. The protocol in this chapter is compatible with purification of the thiolated RNA followed by recoding and sequencing.

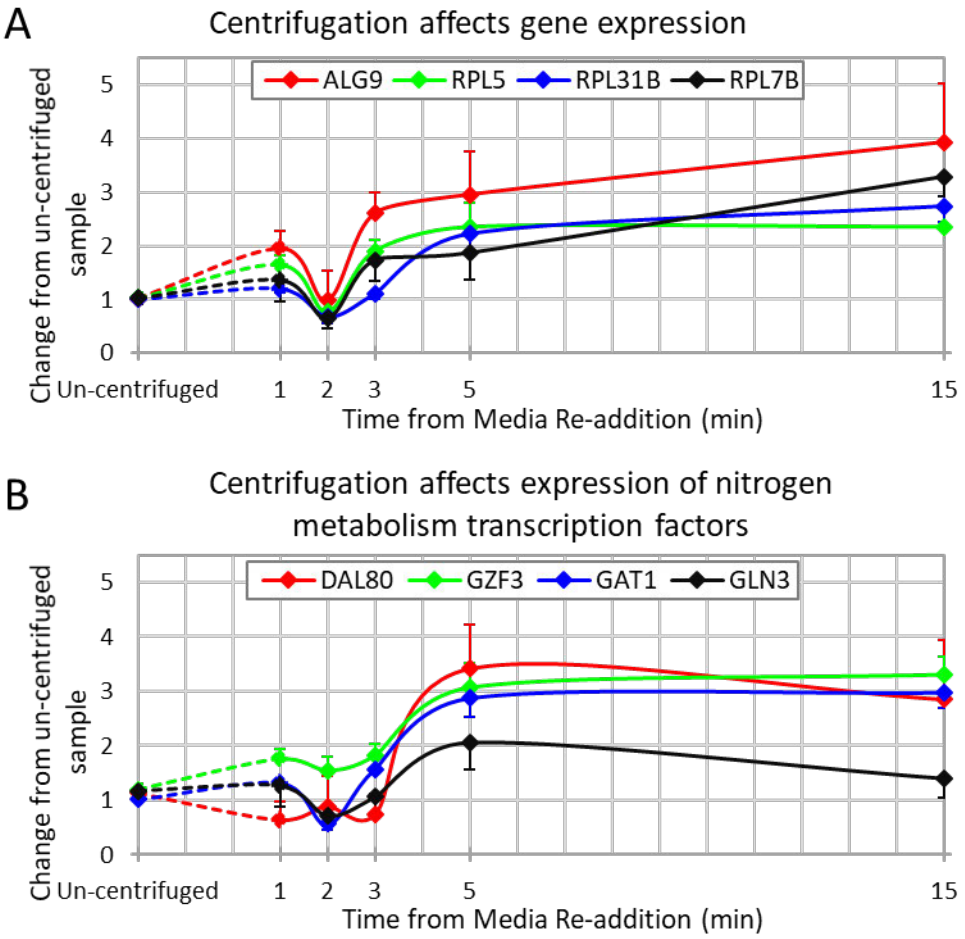


Figure 4: Centrifugation affects gene expression
Thio-labelling for 1 minute, just before centrifugation and 1, 2, 3, 5 and 15 minutes after centrifuging (3,000g for 3 min) and resuspending back in the media it was growing in before. mRNA from the indicated gene detected using RT-qPCR.
A The expression of ribosomal protein genes and *ALG9* clearly dips and recovers **B** Transcription factors involved in nitrogen metabolism change expression. Negative regulators (DAL80 and GZF3) and positive regulators (GAT and GLN3) both increase expression but show signs of returning to basal level by 15 minutes post centrifugation.
2 biological replicates, error bars are standard error

Above all, the primary advantages of the ers4tU protocol over others are that it is sensitive, specific and background is greatly reduced, conferring power and flexibility.

3.6.2 Limitations

The ers4tU protocol has limitations that are important to appreciate. In order to achieve extremely brief thio-labelling the culture has to be grown without

uracil or uridine in the growth medium, therefore, it is necessary to use a URA+ strain. Also, because uracil is limiting, it is conceivable that the addition of any uracil, even in the small amounts recommended, could cause a pulse of transcription. Although this seems unlikely in light of simulations where this is modelled (Figure 3).

In 4-thiouracil, an oxygen that is important in base-pairing is replaced by sulphur, which will have reduced base-pairing potential that could result in errors during transcription and could potentially affect RNA secondary structure. Growth of yeast in the presence of 10 μM 4tU is slightly slowed (P.Figure 1a). Growth recovers with a chase, so much so that any transient reduction is not noticeable in a growth curve; any toxic effects are not permanent. There is a report of rRNA processing disruption in human cells after prolonged growth with >50 μM 4sU (Burger et al., 2013); much longer times and higher concentrations than those used here. A possible reason could be that pseudouridylation of rRNA is inhibited. Poor growth is a problem common to all modified nucleotides and it is hoped that these effects are minor in the few seconds to minutes of thiolation possible here.

As the nucleotide analogue is uracil, transcripts containing many Us could be overrepresented in the nsRNA. This has been tested with the system used here and found not to be significant other than for the U snRNAs (Barrass et al., 2015), however this apparent enrichment is in doubt, see chapter 4.

Polymerases paused on genes throughout the thiolation time will not produce transcripts that can be purified. Strains that transcribe poorly, either because of a mutant phenotype or poor growth conditions such as starvation, are also poor subjects for ers4tU, although the techniques used here will nevertheless improve recovery of newly synthesised RNA compared to other methods. Longer times and increased culture volumes may be necessary in these strains and conditions. Note that uracil is a good source of nitrogen and so this method cannot be used for studies involving nitrogen starvation (data not shown).

3.6.3 Applications of this Method

This technique was used to identify and purify nsRNAs (P.Figure 3) and to estimate the rates at which they approach steady state levels. High throughput sequencing of transcripts isolated in this way can provide details of transcription, RNA processing and RNA decay transcriptome-wide, as in chapter 4. However, useful information can also be obtained for individual genes using RT-qPCR, for example to analyse the kinetics of pre-mRNA splicing (Aslanzadeh et al., 2018).

The ers4tU protocol can be combined with targeted protein depletion, see chapters 6 and 7, to monitor the effect on RNA metabolism as a protein level declines.

The ers4tU protocol is particularly useful for analysis of short-lived RNAs, many of which are so rapidly degraded that they cannot be identified without crippling the degradation machinery. Examples include cryptic unstable transcripts (CUTs), see chapter 4, promoter proximal events such as promoter proximal pausing (Adelman and Lis, 2012) and antisense transcription from a promoter (PROMPTs) (Preker et al., 2008). The ers4tU protocol is exceptional in permitting highly transient RNA species to be captured and analysed under near physiological conditions, a huge advantage over other methods.

3.7 Contribution

Although the fundamentals were established previously (Dolken et al., 2008), the work presented was developed anew from first principles by myself. All data and figures were generated by me and I am the corresponding author of this publication.

All additional data presented in this chapter are also my own, as is any re-interpretation of the results and conclusions from the publication.

3.8 References

Most references are in the publication, a few additional ones are listed here.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Barrass, J.D., and Beggs, J.D. (2019). Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE J. Vis. Exp. e59952.

Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. *16*, 282.

Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. *10*, 1623–1630.

Chevallier, M.R. (1982). Cloning and transcriptional control of a eucaryotic permease gene. Mol. Cell. Biol. *2*, 977–984.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Duffy, E.E., and Simon, M.D. (2009). Enriching s4U-RNA Using Methane Thiosulfonate (MTS) Chemistry. In Current Protocols in Chemical Biology, (John Wiley & Sons, Inc.) *8*, 234-250.

García-Martínez, J., Aranda, A., and Pérez-Ortín, J.E. (2004). Genomic Run-On Evaluates Transcription Rates for All Yeast Genes and Identifies Gene Regulatory Mechanisms. Mol. Cell *15*, 303–313.

Gustilo, E.M., Vendeix, F.A.P., and Agris, P.F. (2008). tRNA's Modifications Bring Order to Gene Expression. Curr. Opin. Microbiol. *11*, 134–140.

Hobro, A.J., and Smith, N.I. (2017). An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. Vib. Spectrosc. *91*, 31–45.

Hokin, L.E., and Hokin, M.R. (1954). Ribonuclease and ribonucleic acid synthesis. J. Histochem. Cytochem. *2*, 395–400.

Marchal, C., Haguenauer-Tsapis, R., and Urban-Grimal, D. (2000). Casein Kinase I-dependent Phosphorylation within a PEST Sequence and Ubiquitination at Nearby Lysines Signal Endocytosis of Yeast Uracil Permease. J. Biol. Chem. *275*, 23608–23614.

Mielcarek, A., and Dołęga, A. (2016). Weak hydrogen bonding interaction S–H···OC studied by FT-IR spectroscopy and DFT calculations. J. Mol. Struct. *1103*, 217–223.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol. Syst. Biol. *7*, 458.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol. Biol. Cell *22*, 2787–2795.

Neymotin, B., Athanasiadou, R., and Gresham, D. (2014). Determination of in vivo RNA kinetics using RATE-seq. RNA *20*, 1645–1652.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. Science *322*, 1851–1854.

Riml, C., Amort, T., Rieder, D., Gasser, C., Lusser, A., and Micura, R. (2017). Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. Angew. Chem. Int. Ed. *56*, 13479–13483.

Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods *15*, 221–225.

Tani, H., and Akimitsu, N. (2012). Genome-wide technology for determining RNA stability in mammalian cells. RNA Biol. *9*, 1233–1238.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived non-coding transcripts in mammals. Genome Res.*5*, 947-56.

Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res. *22*, 2031–2042.

Zhang, J., Smith, K.M., Tackaberry, T., Sun, X., Carpenter, P., Slugoski, M.D., Robins, M.J., Nielsen, L.P.C., Nowak, I., Baldwin, S.A., et al. (2006). Characterization of the Transport Mechanism and Permeant Binding Profile of the Uridine Permease Fui1p of Saccharomyces cerevisiae. J. Biol. Chem. *281*, 28210–28221.

3.9 Reprint

jove Journal of Visualized Experiments www.jove.com

Video Article

Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU)

J. David Barrass[1], Jean D. Beggs[1]

[1]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh

Correspondence to: J. David Barrass at David.Barrass@ed.ac.uk

URL: https://www.jove.com/video/59952
DOI: doi:10.3791/59952

Keywords: Biochemistry, Issue 150, *Saccharomyces cerevisiae*, nascent RNA, newly synthesized, transcription, splicing, RNA processing, RNA degradation, RNA metabolism, pulse chase

Date Published: 8/22/2019

Citation: Barrass, J.D., Beggs, J.D. Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). *J. Vis. Exp.* (150), e59952, doi:10.3791/59952 (2019).

Abstract

The nucleotide analogue, 4-thiouracil (4tU), is readily taken up by cells and incorporated into RNA as it is transcribed in vivo, allowing isolation of the RNA produced during a brief period of labelling. This is done by attaching a biotin moiety to the incorporated thio group and affinity purifying, using streptavidin coated beads. Achieving a good yield of pure, newly synthesized RNA that is free of pre-existing RNA makes shorter labelling times possible and permits increased temporal resolution in kinetic studies. This is a protocol for very specific, high yield purification of newly synthesized RNA. The protocol presented here describes how RNA is extracted from the yeast *Saccharomyces cerevisiae*. However, the protocol for purification of thiolated RNA from total RNA should be effective using RNA from any organism once it has been extracted from the cells. The purified RNA is suitable for analysis by many widely used techniques, such as reverse transcriptase-qPCR, RNA-seq and SLAM-seq. The specificity, sensitivity and flexibility of this technique allow unparalleled insights into RNA metabolism.

Video Link

The video component of this article can be found at https://www.jove.com/video/59952/

Introduction

RNA has a dynamic nature; soon after it is produced much RNA is rapidly processed and degraded. Currently, most studies of RNA metabolism analyze the total cellular RNA, which is mostly fully processed and at steady state level. This level depends on the balance between the rates of transcription, post-transcriptional maturation and degradation. Analysis of the processes that lead to the steady state equilibrium requires specialized techniques to capture very short-lived RNA species.

Metabolic labelling of RNA with nucleotide analogues such as 4-thiouracil (4tU) or 4-thiouridine (4sU) (see Duffy et al.[1] for an excellent review), offers the ability to isolate thio-labelled nascent RNAs and their processing intermediates. However, published protocols involve labelling times of several minutes[2,3], which is slow relative to the rate of production of many transcripts. It takes in the order of one minute to transcribe the average yeast gene, so labelling yeast RNA for less than one minute can be considered extremely short. The extremely rapid and specific 4 thiouracil protocol (ers4tU) maximizes the signal to noise ratio by maximizing 4tU incorporation and minimizing the recovery of unlabeled, pre-existing RNA making very short labelling times possible[4].

The thio-modified base must be imported into the cells rapidly and in sufficient quantity to efficiently label the newly synthesized RNA (nsRNA). To promote this, cells are grown in uracil-free medium, and expression of an appropriate permease helps to boost 4tU or 4sU uptake (see **Table 1** for a list of plasmids that carry suitable permease genes and **Supplementary Figure 1**). 4tU's solubility in sodium hydroxide avoids the need for toxic organic solvents required by other nucleotide analogues. Unfortunately, growing cultures for long periods with thio-modified nucleosides at concentrations greater than 50 µM has been observed to disrupt ribosomes[5]. However, the concentration (10 µM) used here, and the extremely short labelling times, minimize deleterious effects[5] (**Figure 1a**), while still yielding sufficient RNA for analysis.

This technique can be combined with rapid and specific auxin-mediated depletion of a target protein[6,7] (**Figure 2**), referred to as the "β-est AID 4U" protocol, in which β-estradiol regulated expression of the auxin inducible degron (AID) system is combined with 4tU labelling. With the β-est AID 4U approach, a target protein can be depleted and the effect on RNA metabolism closely monitored (**Figure 2**). The timing is critical; it is advisable to view the accompanying video and pay close attention to **Figure 2** and its animated form (see **Supplementary Figure 2**).

Processing and degradation of RNA must be stopped extremely rapidly for accurate time resolution. This is achieved using methanol at low temperature, which fixes the cell contents very rapidly and degrades the cell membrane while preserving the nucleic acid content[8]. The RNA extraction should be efficient and not damage the RNA. Mechanical lysis is effective in the absence of chaotropic agents (often these contain thio groups, so should be avoided). Lithium chloride precipitation of RNA is preferred, as tRNAs are less efficiently precipitated. tRNAs are rapidly transcribed and naturally thiolated[9], so removing tRNAs reduces competition for the biotinylation reagent. If small, highly structured RNAs are of interest, alcohol-based RNA precipitation methods are recommended.

jove Journal of Visualized Experiments www.jove.com

To recover the thiolated RNA, biotin is covalently attached via the thio groups incorporated into the RNA with 4tU. The use of modified biotin, which attaches via a cleavable disulfide bond (e.g., HPDP-biotin (N-[6-(Biotinamido)hexyl]-3´-(2´-pyridyldithio)propionamide,) or MTS-biotin (Methane thiosulfonate)) is recommended as it permits release of the RNA by addition of a reducing agent. The biotinylated RNA is affinity purified on streptavidin coupled to magnetic beads. This protocol is similar to others listed previously[10] but has been intensively optimized to reduce background.

There are two types of thiol-labelling experiment that can be performed, continuous and discontinuous labelling. Each has its own advantages. In continuous labelling the 4tU is added to the culture and samples taken at regular intervals. This type of experiment shows how the RNA is processed and how levels change over time. Examples include comparison of mutant with wild-type experiments and a pulse-chase experiment. The experiments shown in **Figure 3b,c** are of this type. For discontinuous labelling a change is induced into the system and the RNA monitored. Once the change has been induced the culture must be split into several sub-cultures, and at specific times, each one is then thio-labelled for a brief period. One example is β-est AID 4U shown in **Figure 2**[7]. This type of experiment is particularly useful for monitoring the effect of a metabolic change on RNA processing (see **Figure 3d**).

A graphical representation of a thio-labelling experiment is presented in **Figure 4** and **Figure 5**, and a spreadsheet that greatly simplifies the performance of the protocol is available (see **4tU experiment template.xlsx**). As well as this the Supplementary Information contains an extensive troubleshooting guide. For the β-est AID 4U protocol that integrates 4tU labelling with the auxin depletion protocol, see **Figure 2** and **Supplementary Figure 2**. See Barrass et al.[7] for the detailed AID depletion protocol.

Protocol

1. Growth and thio-labelling

NOTE: Time for completion of this section of the protocol is highly variable, depending on cell growth rate. Allow 1 h to prepare the solutions and equipment prior to thio-labelling and 30 min post-labelling to process samples.

1. Ensure the *S. cerevisiae* strain contains a plasmid encoding a permease (**Table 1**) to boost 4tU import into the cell.
   NOTE: Without an importer, labelling for less than 2 min is unlikely to be successful[11] (see **Supplementary Figure 1**). 4tU incorporation is more efficient if growth is in medium without uracil, so the strain must be *URA3*+; several of the plasmids in **Table 1** carry *URA3* as marker. If this protocol is to be combined with β-est AID depletion[7], additional strain modifications are required.
2. Prepare YMM uracil-free medium by adding 6.9 g of yeast nitrogen base without amino acids, 20 g of glucose, and 1.92 g of SCSM single drop-out−ura (**Table of Materials**) to 1 L of water. Autoclave or filter sterilize the growth medium before use.
   NOTE: Filter sterilization is preferred as peptide/sugar complexes produced by autoclaving co-precipitate with the cells in the methanol used in sample collection.
3. Grow yeast in YMM uracil-free medium to an optical density at 600 nm (OD_{600}) of 0.6−0.8. Ensure the culture is in log phase growth and has been for at least two doublings. Growth at 30 °C is normally recommended, but other temperatures may be used, for example, for temperature-sensitive strains.
   NOTE: Depending on the strain, growth conditions and RNA yield, approximately 30 mL sample volume will be needed. This amount will be assumed throughout the protocol. 30 mL of culture is the most that will fit into a 50 mL centrifuge tube with 20 mL of methanol, so is a convenient volume to start optimization. Consider using more sample volume for early time points to increase RNA recovery, up to 2000 mL has been used for slower growing cells at really short labelling times (<1 min).
4. Chill about 50 mL of H_2O on ice. For each sample, add 200 µL of zirconia beads to a 2 mL screw-cap tube and chill on ice. Also put 20 mL of methanol (CAUTION), into 50 mL centrifuge tubes, and place on dry ice (CAUTION). The methanol should be $^1/_3$ to $^2/_3$ the volume of the sample.
   CAUTION: Methanol is toxic by inhalation, contact and consumption. Dispense large volumes in a fume hood, and wear two pairs of gloves, as methanol can penetrate nitrile laboratory gloves. Methanol is highly flammable, keep away from all sources of ignition.
   NOTE: As dry ice can cause cold burns on contact and produces asphixiant gas, use gloves when handling and use in a well-ventilated space.
   NOTE: Adding the beads at this point is easier than after the sample has been added as the tube is dry and when spinning down the cell pellet the beads are also spun clear of the tube thread saving some time. Additionally, this allows the beads to cool before the sample is added.
5. If an *S. pombe* spike is to be added to the culture (rather than later), thaw an aliquot of thiolated *S. pombe* cells on ice and vortex thoroughly, at least 30 s, then add to the culture. If prepared according to the instructions below, one *S. pombe* aliquot is sufficient for 400 mL of culture (enough for twelve 30 mL samples plus a little to allow for errors in handling). If more or less culture is used, adjust the volume of *S. pombe* added to the culture.
   1. Grow 1 L of *S. pombe* culture to OD_{600} to 0.8 exactly as described in the protocol for *S. cerevisiae*.
   2. Thio-label as step 1.7, but for 10 min.
   3. Fix all of the culture using 400 mL of methanol on dry ice, essentially as described in step 1.9.
   4. Pellet the cells by centrifugation at 3000 x *g* for 3 min.
   5. Discard the supernatant and resuspend the cell pellet in 3.3 mL of H_2O.
   6. Split into aliquots of 80 µL each. Store at -80 °C.
   7. Use all of one aliquot for 400 mL culture or 10 µL per 30 mL sample.
      NOTE: Reduce the volume of spike to $^1/_{10}$ if performing RNAseq. Do not reuse aliquots; discard any unused spike. This spike is useful to normalize and compare results across time points and experiments.
6. For discontinuous labelling, induce the required metabolic perturbation (e.g., growth conditions, gene induction or depletion such as β-est AID[7] (**Figure 2** and **Supplementary Figure 2**), then split the culture. Ensure all flasks and media are at the required temperature and, if possible, aerate the medium before adding the culture.

7. Add 4tU to the culture to a concentration of 10 µmM and mix vigorously ($^1/_{10,000}$ of the culture volume of 100 mM 4tU dissolved in 1 M NaOH). Thiol-label for 15 s to 5 min.
   NOTE: Thirty seconds is a good starting point. Thio-labelling for less than 20 s gives more variable results due to difficulties manipulating the culture under time pressure. However, labelling for more than 1 min reduces the temporal resolution of the technique.
8. If a chase experiment is to be performed; allow thio-labelling for 20−30 s then chase by adding $^1/_{200}$ culture volume of 1 M uridine (not thiolated), to a final concentration of 5 mM.
   NOTE: Uridine is preferable to uracil for the chase, as uridine is more water soluble allowing a smaller volume to be added to the culture and so there is less disturbance to the growth of the cells.
9. Take samples of culture at regular intervals (at least 15 s), to the end of the time course. Sampling intervals shorter than this are difficult to perform reliably. Add the sample to the methanol on dry ice prepared in step 1.4. For convenience, add 30 mL of culture to a 50 mL tube containing 20 mL of methanol.
   NOTE: Carbon dioxide dissolves in the methanol when cold; this comes out of solution on addition of the sample and foams vigorously upon mixing—resulting in sample loss. To avoid this, chill the methanol to <-70 °C in a tightly sealed tube until close to the time it is needed, then transfer to dry ice.
10. Seal the tube and mix thoroughly by shaking. Place the samples on ice. Check that none of the samples have frozen; if so, gently warm in the hand, inverting constantly. This is best done in the hand as the sample's temperature can be assessed, it should always feel cold. Place on ice. This is not a pause point; once all the sample is fluid proceed to the next step.
11. Spin at 3000 x *g* for 2 min (at 4 °C if possible) to pellet the cells. Pour off the liquid and resuspend the pellet in at least 1 mL of ice-cold water by gently pipetting up and down.
    NOTE: The residual methanol in the sample pellet aids resuspension.
12. Transfer to 2 mL screw cap tubes as prepared in step 1.4. Spin briefly (e.g., 10 s total time) at >13,000 x *g* to re-pellet the cells, place back on ice and remove liquid.
    NOTE: The cell pellet can be stored at -70 to -80 °C for several months.

2. Preparation of total RNA

NOTE: The time for completion is 90 min.

1. Use diethyl pyrocarbonate (DEPC)-treated solutions to protect the RNA from degradation. Aliquot the solutions using filter pipette tips and wear gloves at all times.
   1. To a solution add $^1/_{1000}$ volume of DEPC and mix by vigorous shaking.
   2. Leave at room temperature (RT) for 24 h, then autoclave.
   3. Solutions with amine groups (such as tris) cannot be DEPC treated. Aliquot the powder and store specially for RNA work. Use previously DEPC treated H_2O to make the solution.
      NOTE: As the thio-group on the RNA is photoactivatable, minimize exposure to UV light from this point on. Storage should be in the dark and incubation is best done in a PCR machine with a lid.
2. If an *S. pombe* spike is to be added to the cell pellet rather than the culture (do not do both), add it now. Thaw an aliquot of thiolated *S. pombe* cells on ice and vortex thoroughly, at least 30 s, before adding to the pellet.
   NOTE: If prepared according to steps 1.5.1−1.5.7, 10 µL of *S. pombe* aliquot is required for one pellet derived from 30 mL of culture.
3. Before putting on the cap, spin very briefly for 1−2 s to ensure that no zirconia beads are trapped between the cap and the tube, which can cause sample and phenol to leak from the tube.
4. Resuspend the cells in 400 µL of acetate EDTA (AE) buffer (50 mM sodium acetate pH 5.3, 10 mM EDTA pH 8.0), by vortexing vigorously. Add 40 µL of 10% (w/v) sodium dodecyl sulfate (SDS). Do not vortex, as SDS will foam.
5. If the β-est AID 4U protocol is to be used, take 40 µL of the cell suspension for protein analysis[7]. Add 40 µL of AE to make the volume back up to 400 µL.
6. Add 800 µL of phenol (CAUTION) at low pH and vortex for 10 s.
   CAUTION: Phenol is toxic and corrosive by inhalation and contact. Always perform procedures involving phenol in a fume hood and wear two pairs of gloves.
7. Lyse the cells in a homogenizer (e.g., **Table of Materials**) for three 2-min bursts at the lowest power setting. Leave the samples on ice for 2 min between pulses of homogenization.
   NOTE: Optimize the conditions if using other homogenizers. Insufficient shaking will result in poor yields, whereas excessive shaking results in apparent higher yield, as determined by absorbance at 260 nm (A_{260}), but the RNA may be degraded. A homogenizer is preferred, but hot phenol RNA purification[12] can be used.
8. Place the lysed sample on dry ice for 5 min, until it solidifies, this reduces genomic DNA carry over into the RNA. Do not freeze for too long as the sample will not thaw. Spin 5 min in microfuge at >13,000 x *g* at RT; do not be tempted to do this at 4 °C, as the sample/phenol mix will remain solid throughout the spin if performed at low temperature.
   NOTE: If the sample is still frozen at the end of the spin, re-spin for another 5 min until the sample has completely thawed.
9. Phenol/chloroform extract then chloroform extract with an equal volume (approximately 600 µL) of phenol:chloroform 5:1 then chloroform (CAUTION). Transfer the top phase to another tube containing phenol:chloroform 5:1 or chloroform. Vortex, then spin for 5 min in a microfuge at RT. Then transfer the top phase to a new 1.5 mL tube.
   CAUTION: Chloroform is toxic by inhalation and contact. Always perform procedures involving chloroform in a fume hood and wear two pairs of gloves.
10. Add a third to half volume (approximately 300 µL) of 10 M LiCl, and mix to precipitate the RNA. The sample should go cloudy immediately but leave for at least 10 min on ice or at 4 °C (do not store below -20 °C as it will freeze), or until the precipitate flocculates.
11. Spin for 5 min at >13,000 x *g* in a microfuge. Remove the fluid, briefly re-spin and remove the dregs. Wash pellet with 300−500 µL of 70% ethanol, spin briefly and remove remaining ethanol.
    NOTE: During these washes keep the pellet on the same side of the tube as the first spin, this way the pellet will not move and break; if it breaks some of the RNA could be lost accidently.

NOTE: Do not dry the pellet; as long as most of the fluid has been removed it will not interfere with subsequent steps. The RNA can also be stored at this stage at -20 °C for a few months or -70 to -80 °C for long-term storage.

12. Re-dissolve the RNA pellet in 90 µL of TE pH 7.0 (10 mM Tris HCl pH 7.0, 1 mM EDTA pH 8.0) by heating at 65 °C with shaking as the RNA pellet can be difficult to re-dissolve. This must be for **no more than 5 min** as RNA degrades at higher temperatures. Check for full RNA solubilization and then transfer to a 0.2 mL tube. Pipette the sample up and down; there should be no "lumps", and the fluid should rise and fall smoothly in the tip. This solution is viscous so the final pipetting motion should be slow.
NOTE: The RNA can be stored at -20 °C in the dark at this stage; this can also be beneficial to RNA solubility.

3. Biotinylation

NOTE: The time for completion is 60 min. The following steps are conveniently done in a strip of tubes with integral caps as they have less tendency to open on vortexing than strips with separate caps.

1. Biotinylate by adding 10 µL ($^1/_{10}$ final volume) of a 5 mM HPDP-biotin solution (MTS-biotin can be used in exactly the same way as HPDP-biotin), to the RNA and mix thoroughly. Preheat the RNA for no more than a few seconds at 65 °C before adding the biotin. Incubate at 65 °C for 15 min to a **maximum of** 30 min in the dark.
NOTE: This heating is required as some HPDP batches precipitate at RT in the RNA sample. A PCR block with a heated lid is ideal for this.
2. Prepare a small resin volume, size exclusion column (**Table of Materials**) to exclude the unincorporated biotin. Remove the bottom tag of the column and loosen the cap, place in a 2 mL centrifuge tube. Spin at 1500 x *g* for 1 min to flush out the buffer Add 0.3 mL of TE gently to the top of the column and spin again. Repeat the wash and spin twice more for a total of 3 washes. Finally transfer the washed column to a fresh 1.5 mL tube.
3. Once the sample incubation (step 3.1) is complete, add the sample to the top of the column. Spin at 1500 x *g* for **2 min.** The biotinylated RNA sample is now in the bottom of the tube.
NOTE: A 1 min spin is not sufficient to elute the entire sample.
4. Add a third to half volume (approximately 40 µL) of 10 M LiCl, mix to re-precipitate the RNA as step 2.10. The sample should go cloudy immediately but leave for at least 5 min on ice or at 4 °C or until the precipitate flocculates; do not store below -20 °C as it will freeze. Centrifuge the sample for 5 min at >13,000 x *g* in a microfuge.
5. Wash with 80% ethanol, ≤1 h rotating. Follow the procedure in step 2.11 to remove as much of the fluid as possible.
NOTE: As HPDP-biotin is very soluble in 80% ethanol, this is an additional purification step.
6. Repeat the 80% ethanol wash to remove as much un-incorporated biotin as possible.
NOTE: The RNA can also be stored at this stage at -20 °C in the dark.

4. Purification of the newly synthesized RNA

NOTE: The time for completion is 2 h.

1. Re-dissolve the RNA in 200 µL of DEPC-treated H_2O (65 °C incubation can be used, similar to the procedure in step 2.12).
2. Measure the RNA concentration at A_{260} using a spectrophotometer; the sample may have to be diluted $^1/_{10}$ to get it within the linear range of the spectrophotometer. Vortex this dilution for at least 10 s to ensure the viscous RNA is evenly dissolved.
NOTE: The efficiency of biotinylation can be assessed by dot blot[13] if required.
3. Add equal amounts of RNA to a fresh tube and make up to 200 µL in DEPC-treated H_2O. Use all of the sample with the lowest RNA concentration and use an appropriate volume of the other samples to have a similar amount of RNA for each.
NOTE: The spreadsheet **4tU experiment template.xlsx** has a form to aid this calculation.
4. When the sample is at RT, add 25 µL of 10 x NaTM buffer (0.1 M Tris HCl pH 7.0, 2 M NaCl, 250 mM $MgCl_2$), 25 µL of 1 M NaPi pH 6.8 (0.5 M NaH_2PO_4 0.5 M Na_2HPO_4), and 2.5 µL of 10% SDS. Mix thoroughly and spin gently (<30 s; approximately 100 x *g*).
NOTE: To avoid precipitation of the SDS and salts, the samples must be kept at RT throughout the following procedures up to step 4.13.
5. Make the bead buffer containing 1x NaTM buffer, 0.1 M NaPi, and 0.1% SDS, 2 mL per sample. Add the required amount of H_2O first and the SDS last. This must be made fresh each time as a precipitate forms after 24 h.
NOTE: To avoid the formation of precipitates, the bead buffer must be kept at RT throughout the following procedures. Do not DEPC treat or autoclave.
6. Add 50 µL of streptavidin beads to a low retention 1.5 mL tube. Place the tube on the magnetic rack, wait for the beads to settle and then remove the fluid
7. Wash the streptavidin beads.
    1. Add 200 µL of bead buffer and vortex until the bead pellet is fully resuspended. Usually 3−5 s is all that is required. For washes before the RNA sample is added it is sufficient to turn the tubes round so the beads travel across the tube to the other side. Then turn the tubes back to the original side so the beads travel across the tube once again.
    2. Spin the tube at low speed (approximately 100 x *g*) for a maximum of 5 s to spin down the fluid, but not the beads.
    3. Place in the magnetic rack to allow the beads to be captured by the magnet.
    4. Remove the fluid by aspiration for a small number of samples; pour off the liquid if many samples.
    NOTE: With a large number of samples, removing all the fluid purely by aspiration can be problematic, as the beads in the first sample may be dried out before the last sample is finished, this increases background. Washing can be expedited by pouring off the fluid from all samples at once whilst on the magnet. They should be left a little longer on the magnet before pouring and the small amount of fluid that remains has to be aspirated away but, overall, it means less time on the magnet and without fluid. In this way, it is possible to do 24 or more extractions quickly.
8. Block with 200 µL bead buffer, 10 µL 20 mg/mL glycogen, and 2.5 µL 5 mg/mL tRNA, 20 min rotating end over end at moderate speed at RT. The rotation is to keep the beads in suspension. Once the blocking is complete remove the fluid as steps 4.7.2−4.7.4 and wash again, as the steps in section 4.7.
9. Resuspend the beads in the sample. Incubate at RT with rotation for 30 min.

jove Journal of Visualized Experiments www.jove.com

10. During the incubation, prepare a fresh 1.5 mL tube for each sample. Add $^{1}/_{10}$ volume (approximately 10 µL) of 3 M sodium acetate pH 5.3 and 20 µg of glycogen, and spin at approximately 100 x *g* for 3 s. Store in a rack until needed.
11. Remove the unbound RNA from the beads, as steps 4.7.2−4.7.4. The unbound RNA can be persevered in a fresh tube, but the salts and SDS make it very difficult to purify. Then wash the beads, as section 4.7 with vortexing, for a minimum of 3 to a maximum of 5 times.
12. After the final wash take special care to aspirate all the liquid; return to each tube and aspirate the dregs of the buffer once more.
13. To elute the RNA, add 50 µL of freshly prepared 0.7 M β-mercaptoethanol (βME) to the beads ($^{1}/_{20}$ dilution of the commercially supplied stock solution). Vortex and spin briefly, as steps 4.7.1 and 4.7.2. Place the slurry in the magnetic rack and pipette the RNA containing solution into the 1.5 mL centrifuge tube prepared in step 4.10.
14. Elute once more as step 4.13 to recover residual RNA from the beads and add the eluted sample to the tube containing the first elution from these beads.
15. Remove residual beads from the eluted RNA by placing the sample back in the magnetic rack and transferring the fluid to a fresh, low binding 0.5 mL centrifuge tube.
16. Mix the sample and then precipitate the nsRNA by adding 2.5x volumes (280 µL) of ethanol and mix once more. Leave for 1 h to overnight at -20 °C. Spin in a pre-chilled centrifuge (4 °C) for 20 min at the maximum speed (at least 13,000 x *g*).
17. Wash thoroughly with 200 µL of 70% ethanol at -20 °C. As residual βME will inhibit downstream applications, spin at every step to remove as much of the dregs as possible; at the end the sample should not smell of βME.
18. Re-dissolve in 10−20 µL of DEPC-treated 1x TE with the equivalent of 0.005 µL RNase inhibitor.
    NOTE: All subsequent stages should be performed on ice.
19. Measure the RNA concentration and purity.
    1. Measure the A_{260} and A_{225} in a low sample volume spectrophotometer.
       NOTE: An absorbance maximum near λ = 225 nm is from an unavoidable contaminant from the beads. In the absence of RNA the signal from the contaminant declines to 35% at λ = 260 nm. Therefore, the actual amount of RNA is approximated by the formula: $(A_{260}-(A_{225}*0.35))*40$ ng/µL.
    2. Alternatively, analyze the sample on a micro-fluidics electrophoresis system such as a bioanalyzer.
       NOTE: This analysis is preferable to using a spectrophotometer as RNA integrity can be assessed, the contaminant does not interfere with the quantitation and less sample is required.

20. Analyze the nsRNA.
    NOTE: For example, specific RNAs can be quantified by standard reverse transcriptase qPCR techniques. RNA prepared this way is compatible with library preparation for RNA-seq. Removal of rRNA is not necessary for labelling times of less than 5 min. Recoding SLAMseq[14] can also be performed on this RNA.

Representative Results

Typical yields for nsRNA recovered using this ers4tU protocol are displayed in **Figure 1b**, this has been produced by a bioanalyzer and the trace shows yield of RNA versus size (nucleotides [nt]). Note, in both the bioanalyzer trace and the inset graph, that RNA recovery from time point 0 is a very small portion of that recovered from longer time points - approximately 0.3 µg of RNA recovered from approximately 10^9 cells compared with over twice as much after just 30 s of labelling (0.8 µg of nsRNA) from the same number of cells. RNA recovery at 15 s is more variable as small differences in performing the sampling have a proportionately larger effect on RNA recovery. In the bioanalyzer trace, rRNA precursors can be seen as a peak near 1000 nt and a doublet of peaks at 1700−1800 nt. The abundance of these intermediates increases as thiolation continues.

Thio-labelling was used to quantify splicing of the *ACT1* transcript (**Figure 3**). Thiolation was performed and samples taken at 15 s intervals from the start of thio-labelling and the processing of *ACT1* RNA monitored (**Figure 3a,b**). As can be seen, pre-mRNA is generated (by transcription), and lariats (by the first step of splicing from pre-mRNA), even after just 15 s of labelling. After about 45 s to 1 min, the amounts of lariats and pre-mRNA reach equilibrium with as much of these RNA species being created by transcription as are processed away by splicing.

To produce the data shown in **Figure 3c** the strain was pulsed with 4tU for 25 s and then chased with uridine. The generation of pre-mRNA and lariats reaches a maximum at 1 minute. This compares well with **Figure 3b**; the maximum being achieved after 45 s to reach equilibrium plus the 25 s of the labelling. After the peak, the levels decline as the thio-labelled RNAs are chased through the splicing process.

Figure 3d shows depletion of a protein splicing factor and its effect on RNA metabolism, using the β-est AID 4U system[6,7]. Here, Prp16p is reduced from near physiological levels to 5% of this level after 25 min of depletion. Prp16p is an essential splicing factor for the second step of splicing[15]. Lariats are removed during the second step of splicing (**Figure 3a**), but here they increase above the level of pre-mRNA as Prp16 becomes limiting. At later depletion times, other factors become limiting due to secondary effects, so that levels of lariat decrease, and pre-mRNA levels rise. The level of spliced mRNA declines.

jove Journal of Visualized Experiments www.jove.com

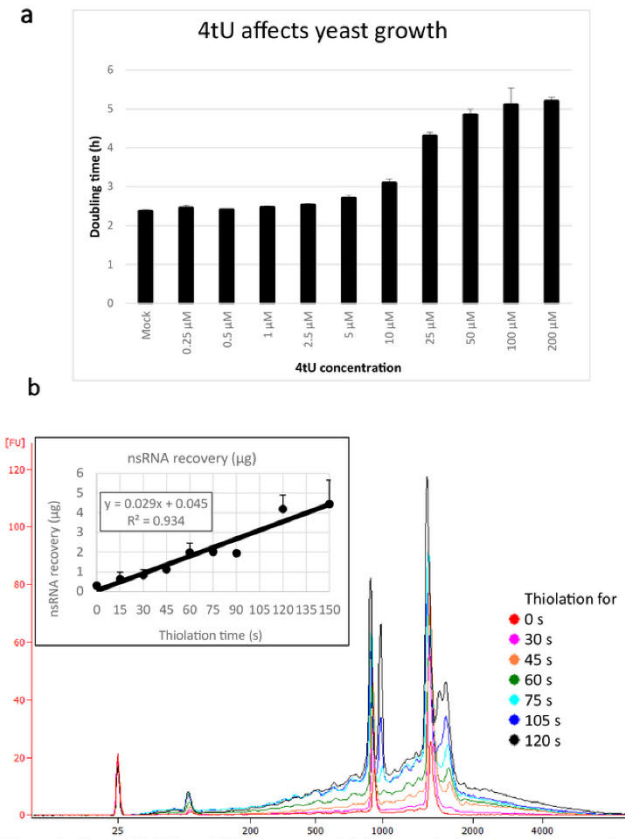


Figure 1: Growth in 4tU and RNA recovery. (**a**) 4-thiouracil affects growth. Increasing the concentration of 4tU in YMM drop-out growth medium without uracil increases the doubling time of *S. cerevisiae* (BY4741) carrying the p4FuiΔPEST plasmid. Growth of four replicate cultures was monitored at 30 °C in a Tecan Infinite Pro 200. All cultures were in log phase throughout, with OD_{600} between 0.1 and 0.6. Mock is a control culture with an equivalent amount of NaOH added, which does not by itself change the growth rate. This graph demonstrates that thio-labelling is compromise between rapid labelling and damage to the cell. Error bars are standard error of 4 replicates. (**b**) nsRNA yield increases linearly from about 15 s of labelling. The main figure shows the bioanalyzer traces of purified, nsRNA from 0 (not thiolated) to 2 min after addition of 4tU at 15 s intervals. Note that the 15 s sample is not shown, as it was indistinguishable from the unlabelled sample. The two large peaks correspond to ribosomal RNAs (rRNAs). The rRNA precursors and intermediates are visible as several peaks at greater molecular weight than mature rRNAs. The recovery of these precursors and intermediates increases with time. Results from one representative experiment are shown. The inset graph shows the recovery of nsRNA with increasing incubation with 4tU. The yield of nsRNA increases with increasing time of growth with 4tU. The recovery is remarkably linear (R^2 = 0.934) throughout the timescale of this experiment and shows a slight increase over background even at 15 s labelling with 4tU even though not distinguishable from the unlabelled sample by eye from the bioanalyzer trace. Error bars show standard error for three biological replicates. Please click here to view a larger version of this figure.

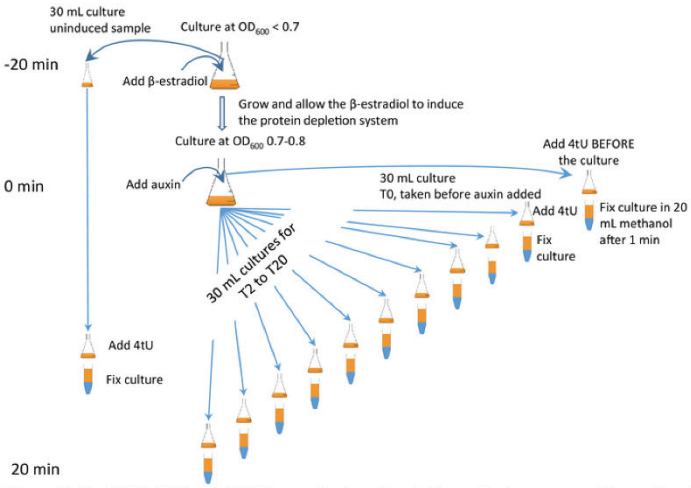


Figure 2: β-est AID 4U β-est AID 4U graphical protocol. A graphical summary of the protocol of the β-est AID 4U protocol. β-estradiol (β-est) promotes the expression of the auxin inducible degron (AID) system which in turn depletes an AID* tagged target protein, refer to Barrass et al.[7] for a detailed protocol. In this case, degron system expression is initiated 25 min before protein degradation commences and thiolation at each time point is for 1 min. Samples are taken before induction and every 2 minutes during depletion. An animated version appears in the **Supplementary Figure 2**. Please click here to view a larger version of this figure.

jove Journal of Visualized Experiments www.jove.com

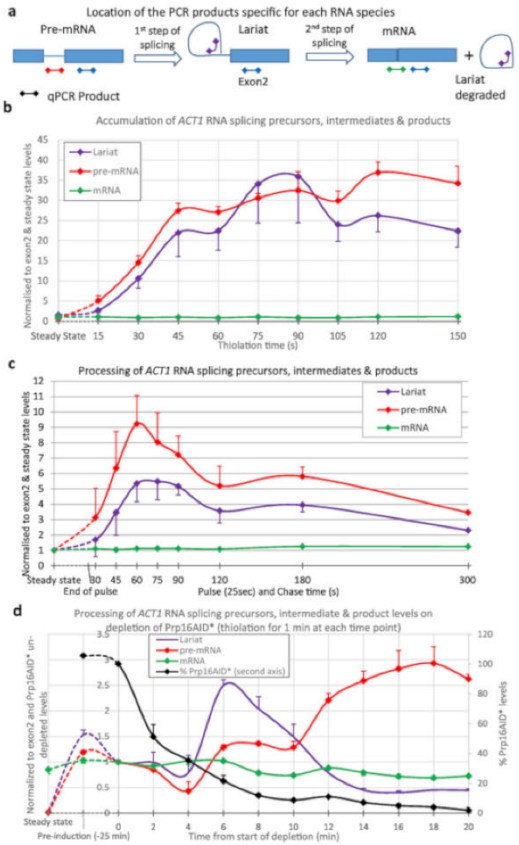


Figure 3: Precursors and intermediates of *ACT1* RNA splicing. Splicing of *ACT1* pre-mRNA transcripts was monitored by quantitative reverse transcription PCR[16]. The levels of *ACT1* precursor (pre-mRNA), intermediate lariat-exon2 (Lariat) and spliced product (mRNA) are shown normalized against the level of *ACT1* Exon2 and steady state levels of these RNAs. (**a**) Location of qPCR products on the ACT1 transcript. Schematic of the locations of the qPCR products used to assay the levels of precursors, intermediates and products of the splicing reaction of the *ACT1* transcripts[16]. Exons are represented by boxes, intron as a line and the qPCR products as lines with diamonds at either end, the color matches those used in the graphs. The pre-mRNA PCR is specific for pre-mRNA and not any intermediates of splicing as this product crosses the branch point which is disrupted after the first step of splicing. Lariat PCR will detect the product of the first step of splicing and the excised lariat produced after the second. The mRNA PCR is specific for the product of splicing, mRNA. Results from the exon PCR (present in all precursors, intermediates and products, except the excised lariat) is not shown in the graphs as this was used to normalize the data and is therefore always equal to 1. (**b**) Continuous thiolabelling. The amount of pre-mRNA increases with time as 4tU is incorporated by transcription and, after a short delay, splicing converts it to lariat-exon2 intermediate and spliced products. The levels of these pre-mRNA and lariat species are detectable above background after as little as 15 s of growth with 4tU and reach a maximum after approximately 45 s of continuous labelling with 4tU, at which point their production is balanced by conversion to spliced mRNA and/or degradation. Values are normalized to their steady state (left-most point of the graph), and exon 2 levels to show their appearance and processing in comparison to transcription of exon 2. As RNA splicing of *ACT1* is largely co-transcriptional[4,17] spliced mRNA rapidly becomes the most abundant species, its level is similar to that of exon 2. Standard error of three biological replicates, each assayed in triplicate. (**c**) Pulse/chase. Thiolation pulse of 25 seconds followed by chase with uridine. Compared to the steady state levels of these RNAs (left-most point), they are initially very abundant in the newly synthesized pool. The levels gradually decline as they are processed into mRNA (or degraded), approaching levels very similar to steady state levels by 5 min. Standard error of three biological replicates, each assayed in triplicate. (**d**) nsRNA and protein depletion. Splicing of *ACT1* pre-mRNA transcripts monitored by quantitative reverse transcription PCR as in panel (**a**) upon depletion of the Prp16 protein using the auxin degron system as described in **Figure 2**. The Prp16 protein levels are also displayed in the graph plotted against the second Y-axis as percentage of levels prior to auxin depletion. Prp16 is a vital component of the spliceosome, particularly important for the second step of splicing shown in panel (**a**), after which lariats are degraded. When this step becomes limiting lariats accumulate initially. At later time points splicing fails completely, lariats are no longer produced and pre-mRNA levels rise. Error bars are standard error of three biological replicates, each assayed in triplicate. Please click here to view a larger version of this figure.

jove Journal of Visualized Experiments www.jove.com

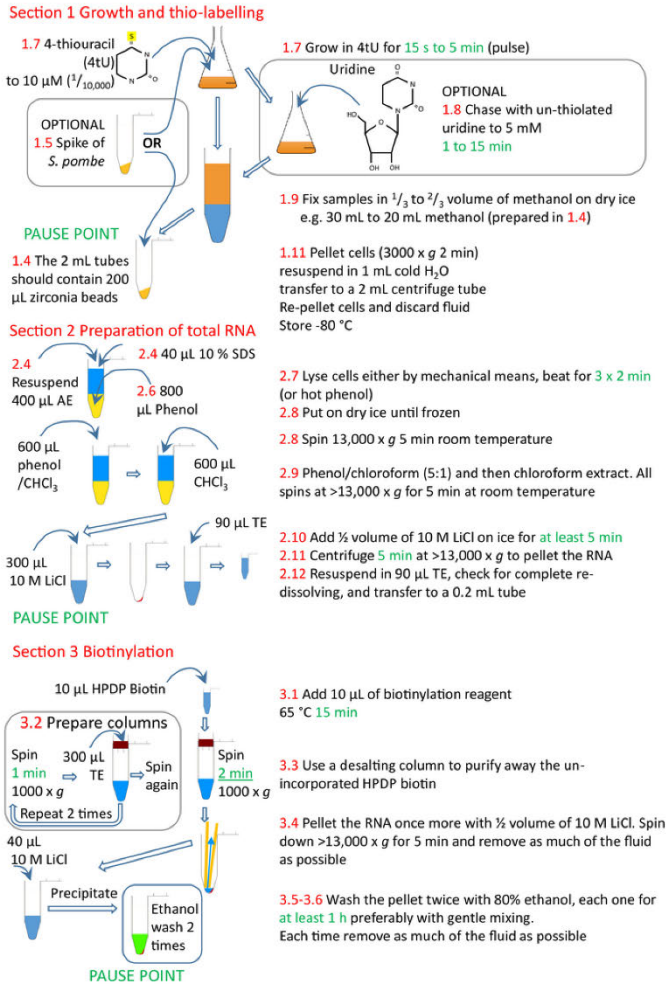


Figure 4: Graphical summary of the protocol sections 1 to 3. The cells are thiolated with 4tU and allowed to grow to incorporate the modified nucleotide into the RNA. A thiolated *S. pombe* spike can be added to allow normalization across time points and experiments. The pulse of 4tU can be chased using un-thiolated uridine. Labelling can either be performed continuously from 4tU addition or from a change to growth conditions, the culture split and 4tU added to cultures at increasing times from the growth condition change, but each labelling only for a brief time. The cells are collected, and RNA prepared from the cells, preferably using a homogenizer and phenol-based methods. The RNA is biotinylated and then the biotinylated RNA purified from unincorporated biotin using a size exclusion column. The nsRNA is now ready for purification with streptavidin beads (section 4, **Figure 5**). Numbers in red correspond to the step numbers in the protocol. Please click here to view a larger version of this figure.

jove Journal of Visualized Experiments www.jove.com

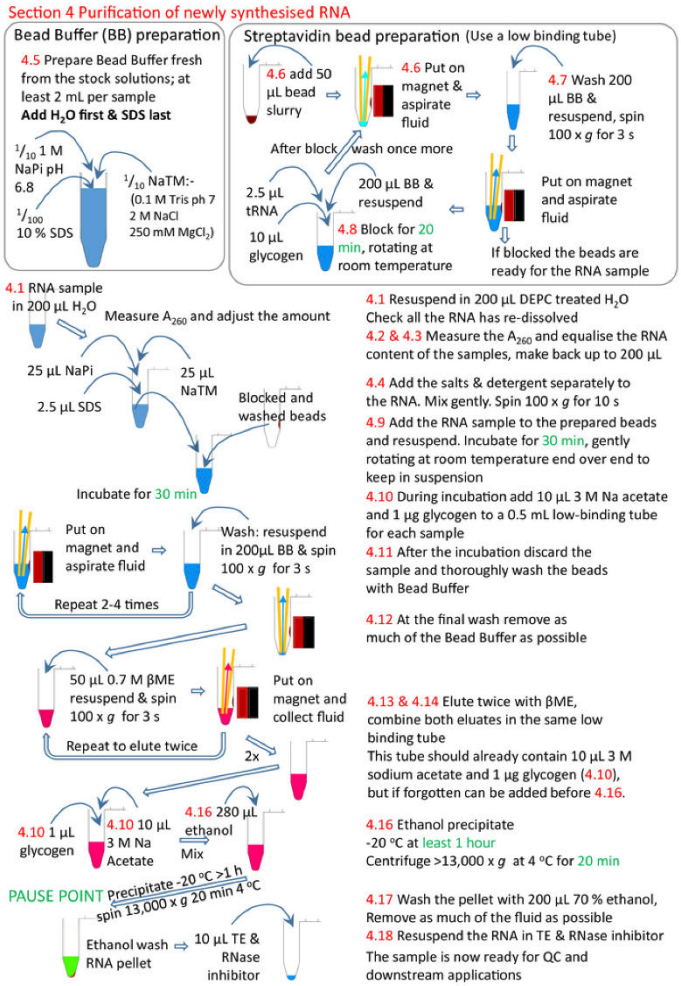


Figure 5: Graphical summary of the protocol section 4. Following on from sections 1 to 3 (**Figure 4**), the streptavidin beads are blocked and the biotinylated RNA sample added to the prepared beads. The biotinylated RNA binds to the streptavidin beads and the un-biotinylated RNA removed and washed. The biotinylated RNA is eluted from the beads using βME and precipitated ready for further research. Numbers in red correspond to the step numbers in the protocol. Please click here to view a larger version of this figure.

Supplementary Figure 1: Improvement of nsRNA recovery from yeast cells with and without additional copies of the importer at 1 and 3 minutes of thio-labelling. Note that *Fui1* is the yeast's own promoter expressed from a 2 µm plasmid. The genomic copy of this gene is present in both of these strains. Please click here to download this file.

Supplementary Figure 2: Animated version of the β-est AID 4U β-est AID 4U graphical protocol. Please click here to download this file.

Supplementary File 1: 4tU_experiment_template.xltx. Please click here to download this file.

Plasmid Name	Importer/permease	Marker	Comment	
p4Fui	*S. cerevisiae* Fui1	*URA3*	Fui1 imports Uracil and Uridine, making it ideal for pulse/chase experiments.	
pAT2	*S. cerevisiae* Fui1	*LEU2*		
p4Fui-ΔPEST	*S. cerevisiae* Fui1	*URA3*	The PEST motif of Fui1 has been deactivated, so the permease is not degraded when there is sufficient intracellular uracil for the cell's needs. Works well in labelling experiments and improves pulse/chase performance.	
p4Fur	*S. cerevisiae* Fur4	*URA3*	Uracil permease	
YEpEBI311	*H. Sapiens ENT1*	*LEU2*	Miller et al.[11]. Also contains an HSV thymidine kinase gene.	
	(equilibrative nucleoside transporter)			
All plasmids are 2 µm based. All p4 plasmids and pAT are based on the pRS[16] series of plasmids. *FUI1* and *FUR4* are expressed from their own, endogenous promoters.				

Table 1: Plasmids used with this protocol.

Discussion

This article presents a protocol for extremely rapid and specific 4tU labelling, for recovery of nascent, newly synthesized RNA from *S. cerevisiae* after as little as 15 s of labelling, with very low contamination by unlabeled RNA.

The user should always take care to maintain the integrity of the RNA by use of cold temperatures and DEPC-treated reagents. Streptavidin bead purification is generally reliable; however, the bead buffer is difficult to handle; it must be made freshly, with its components added in the right order, and not chilled or autoclaved. Common failings include the RNA being incompletely dissolved after the precipitation steps, and so being either not biotinylated or otherwise lost during the processing steps. There is extensive troubleshooting help in the supplementary material.

There are some limitations to be aware of in ers4tU. One already mentioned is that 4tU slows growth of the yeast (**Figure 1a**). Apart from endogenously thiolated RNAs[9], only RNAs that have been transcribed during the labelling period can be purified by this method. Polymerases paused on genes throughout the thiolation time will not produce thiolated transcripts that can be purified, although transcripts that are partially labelled due to polymerases entering or leaving a paused state during thiolation can be recovered. Strains that transcribe poorly, either because of mutation or growth conditions, produce little nsRNA, although the techniques used here will nevertheless improve recovery of nsRNA compared to other methods. Longer times and increased culture volumes may be necessary in these strains and conditions. Note that uracil is a good source of nitrogen and so this method should be trialed before being used for studies involving nitrogen starvation.

The ers4tU protocol is particularly useful for analysis of short-lived RNAs, many of which are so rapidly degraded that they cannot be identified without crippling the degradation machinery. Examples include cryptic unstable transcripts (CUTs)[4], and short transcripts produced by premature termination or promoter proximal pausing[18] and antisense transcription "upstream" from a promoter (PROMPTs)[19]. The intermediates produced during processing of stable RNA species are also transient but can be enriched using ers4tU transcription[4]. The ers4tU protocol is therefore exceptional in permitting highly transient RNA species to be analyzed and captured under near physiological conditions, which is a huge advantage over other methods. This technique has been used to study transcription and downstream RNA processing kinetics in RNA polymerase mutants that elongate faster or slower than normal[20].

Thiolation is also compatible with RNA-seq and SLAM-seq[21], allowing all RNA produced within a very short time window to be characterized in exquisite detail.

Disclosures

The authors have nothing to disclose.

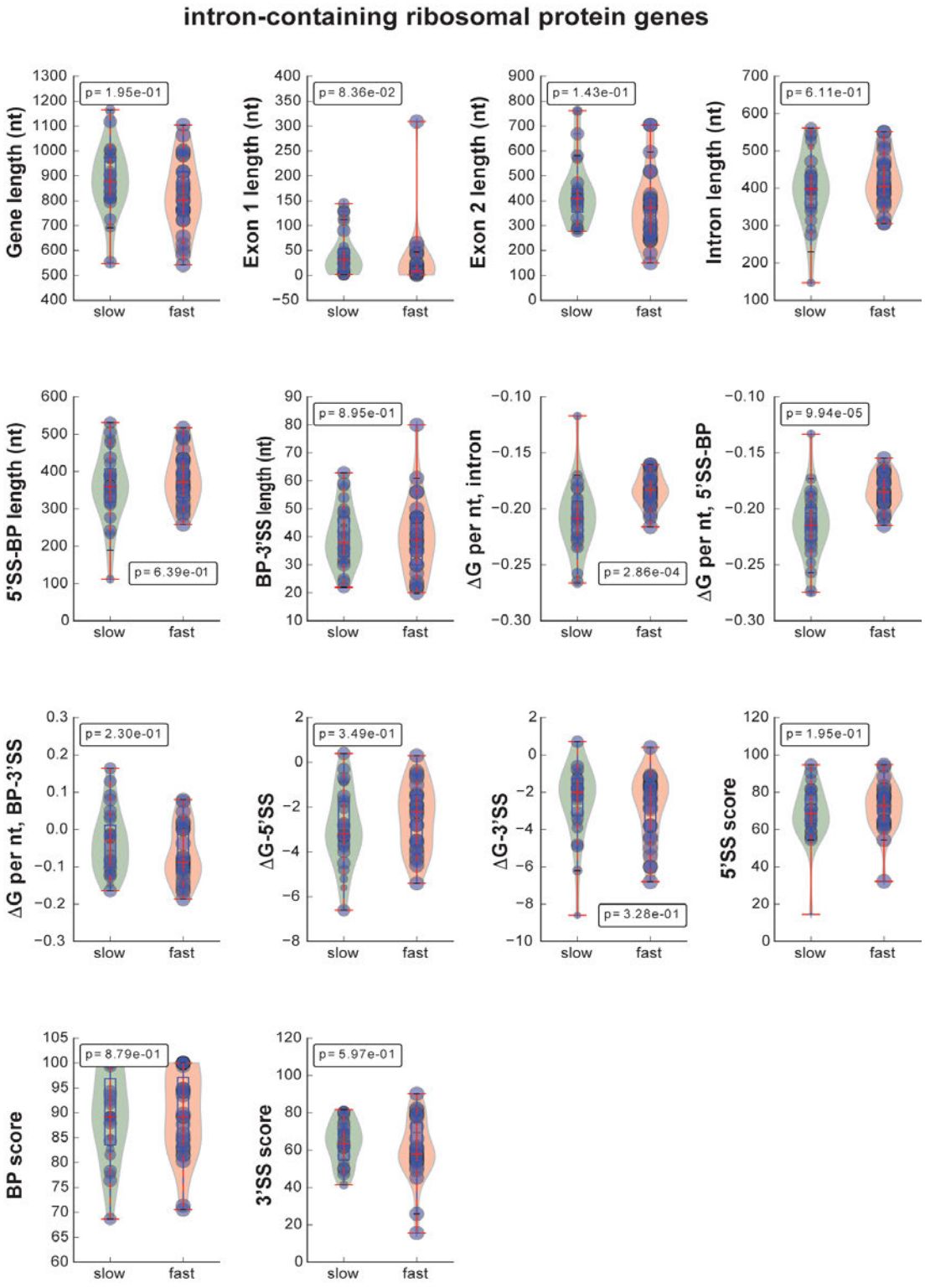
Acknowledgments

This work was supported by Wellcome funding to JB [104648]. Work in the Wellcome Centre for Cell Biology is supported by Wellcome core funding [092076]. The authors acknowledge members of the lab for their help: Bella Maudlin, Emanuela Sani, Susanna De Lucas-Arias and Shiney George. The authors would also like to thank Patrick Cramer for the plasmid YEpEBI311[11].

References

1. Duffy, E. E., Schofield, J. A., Simon, M. D. Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. *Wiley Interdisciplinary Reviews: RNA.* **10** (1), e1513 (2018).
2. Windhager, L. et al. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Research.* **22**, 2031-2042 (2012).
3. Baptista, T., Devys, D. Saccharomyces cerevisiae Metabolic Labeling with 4-thiouracil and the Quantification of Newly Synthesized mRNA As a Proxy for RNA Polymerase II Activity. *Journal of Visualized Experiments.* (140), e57982 (2018).
4. Barrass, J. D. et al. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biology.* **16**, 282 (2015).
5. Burger, K. et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biology.* **10**, 1623-1630 (2013).
6. Mendoza-Ochoa, G. I. et al. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast (Chichester England).* **36** (1), 75-81 (2018).
7. Barrass, J. D., Mendoza-Ochoa, G. I., Maudlin, I. E., Sani, E., Beggs, J. D. Tuning degradation to achieve specific and efficient protein depletion. *Journal of Visualized Experiments.* In press (2019).
8. Hobro, A. J., Smith, N. I. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vibrational Spectroscopy.* **91**, 31-45 (2017).
9. Gustilo, E. M., Vendeix, F. A. P., Agris, P. F. tRNA's Modifications Bring Order to Gene Expression. *Current Opinion in Microbiology.* **11**, 134-140 (2008).
10. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA.* **14**, 1959-1972 (2008).
11. Miller, C. et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular Systems Biology.* **7**, 458 (2011).
12. Schmitt, M. E., Brown, T. A., Trumpower, B. L. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Research.* **18**, 3091-3092 (1990).
13. Rädle, B. et al. Metabolic Labeling of Newly Transcribed RNA for High Resolution Gene Expression Profiling of RNA Synthesis, Processing and Decay in Cell Culture. *Journal of Visualized Experiments.* (78), e50195 (2013).
14. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods.* **14**, 1198-1204 (2017).
15. Ohrt, T. et al. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA.* **19**, 902-915 (2013).
16. Alexander, R. D. et al. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. *RNA.* **16**, 2570-2580 (2010).
17. Wallace, E. W. J., Beggs, J. D. Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. *RNA.* **23**, 601-610 (2017).
18. Adelman, K., Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics.* **13**, 720-731 (2012).
19. Preker, P. et al. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science.* **322**, 1851-1854 (2008).
20. Aslanzadeh, V., Huang, Y., Sanguinetti, G., Beggs, J. D. Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Research.* **28**, 203-213 (2018).
21. Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C., Simon, M. D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nature Methods.* **15**, 221-225 (2018).

4 An Analysis of nascent RNA



Violin plots of factors that might speed or slow splicing, see the publication for details. None of these intronic parameters affected splicing speed.

4.1 Research Article

This chapter is based on the peer-reviewed publication: Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., Granneman, S., 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labelling. Genome Biology 16, 282.

Figures from this publication will be referred to as P.Figure followed by the number.

An analysis of newly synthesised RNA (nsRNA), performed using newer version of the method in the previous chapter

4.2 Aim

To identify and analyse transient RNAs, such as unstable transcripts and processing intermediates in snoRNA maturation and the splicing reaction.

4.3 Experiments

The thio-labelling procedure was almost as described in chapter 3 (Barrass and Beggs, 2019). The RNA extraction was an earlier variant, using “hot phenol” rather than lysis by grinding in a “Beadbeater” (Mini-Beadbeater-24, biospec.com). A standard ethanol precipitation was used to concentrate the RNA, rather than LiCl used in the later protocol, resulting in large amounts of tRNAs biotinylated and purified (as some tRNAs are naturally thiolated). Fifty times more 4tU was also used. As this procedure was far less efficient, larger samples were required, resulting in poor time resolution compared with that achieved later. In this experiment 1.5, 2.5 and 5 minutes of thio-labelling was used. The RNA prepared was subject to high-throughput sequencing with validation of some of the results undertaken by RT-qPCR, as described in the introductory chapter.

4.3.1 nsRNA Recovery

It is interesting to compare RNA recovery from the publication in this chapter with the recovery in the JoVE publication in the previous chapter (chapter 3 Figure 2). Although the protocol was not as optimised the RNA purification is

still well above background level (time 0) and shows impressive linearity of RNA recovery with labelling time.

Two unlabelled samples were also sequenced to provide a background value, <9.1 % of the signal; not an unreasonable value (see chapter 3), and one which does not preclude useful analysis of nsRNA. Transcripts containing more uracil might be more efficiently purified. This was tested for and found not to bias the number of reads (P. Figure S2 and P.Table S2), except the snRNAs, U1 to U6, which are enriched. These are unusually U rich (hence their name), at 34 % U. However, this class also includes U3, one of the most abundant transcripts in the cell. Both copies of this gene have an intron so the increase could be, at least partially, due to their introns being detected in the nsRNA. The number of Us present also correlates with number of reads in the steady state RNA sample, which has not been purified using thiouracil and so cannot have been enriched. The correlations are probably a statistical anomaly due to their high abundance and small number of genes that can be tested.

4.3.2 Thio-labelling Captures Transient RNAs

From P.Figure 2A, several classes of transcripts show considerable enrichment in early thio-labelling time points versus steady state RNA (total RNA). Note that this is numbers of genes significantly enriched, not their degree of enrichment. Intron-containing transcripts and snoRNAs are highly enriched but decline rapidly to only a few more than steady state RNA levels by 5 minutes of thio-labelling. This implies rapid processing/degradation. Anti-sense RNAs and the CUTs, SUTs and XUTs are enriched, and many stay enriched at 5 minutes. For this to be the case these RNAs must persist over the course of the thio-labelling; if they were degraded so rapidly that essentially all of the transcripts in steady state were newly transcribed RNAs, they would not be enriched over steady state in the nsRNA fraction. The tRNAs are always enriched and this does not decline at longer times as the other categories do, the most likely explanation being endogenous thio-labelling (see chapter 3). 26 tRNAs were identified as thio-modified in this study, see P.Figure S1 for these tRNAs.

4.3.3 Processing snoRNAs and rRNA

The processing of rRNAs could not be properly assessed (P.Figure 2B) as, unlike the other RNAs analysed, rRNAs were depleted from the steady state RNA so this sample could not be used as a control. The processing does seem slower than expected, precursors still very apparent even in the 5 minute sample. This could be related to rRNA processing defects possibly due to problems with pseudouridylation of thio-modified uracils in rRNA.

The snoRNAs (P.Figure 2C, D and E), on the other hand, produced some of the most compelling evidence of processing, this can be followed through the time course in the figures. The processing of *snR13* matches the *rrp6* deletion strain in that 3' RNA sequences are retained (P.Figure 2C). The downstream RNA normally removed by Rrp6p is retained in the rrp6Δ strain, so the unprocessed RNA accumulates. The peak just after the promoter of the polycistronic cluster (D) is intriguing and not seen in the steady state RNA. This peak declines as a proportion of the reads through the time course implying that either it is the product of a processing event that leads to its complete elimination in total RNA or that there is abortive transcription from this promoter. A strong possibility is that this is the 5’ stem loop common to s. cerevisiae snoRNAs that is normally cleaved off by the exonuclease Rnt1p (Chanfreau et al., 1998). Rnt1 is co-transcriptionally recruited by the 3’ end processing factor NNS (Nrd1/Nab3/Sen1) complex at the 3’ end of the transcript (Grzechnik et al., 2018). This would therefore be a post-transcriptional event, the cleaved 5’ end of the RNA presumably persisting for a time. Shorter time points would be needed to evaluate the kinetics of this processing event.

One snoRNA, *snR78*, is abundant in steady state, but is poorly represented in nsRNA, it is not clear how this situation arises, but it is probably more stable than other snoRNAs once successfully processed.

4.3.4 Unstable Transcripts

Kinetics of degradation of CUTs and SUTs can also be established from the data. As stated previously they are stable, compared to the introns of pre-mRNAs. In the thio-labelled samples there is little to distinguish CUTs from SUTs at the level of individual transcripts but, as a group, the CUTs are less abundant in steady state. Degradation does not follow first order kinetics, so presumably it is the combination of several steps.

With only three time points (P.Figure S3) it is more difficult than implied in the paper to determine the kinetics of unstable transcript degradation. SUTs do generally persist longer than CUTs and this could be due to their longer length, but this is likely to be a multifactorial process.

4.3.5 mRNA Splicing Kinetics

The RNA-seq data set was filtered to retain only transcripts with one intron, without an intronic snoRNA and with sequencing data of >10 fragments per kilobase per million reads (FPKM). This left 187 transcripts. A probabilistic model was applied to the nsRNA sequencing data. This used the unambiguous reads (those across junctions broken or formed by splicing) to assign probabilities, spliced versus unspliced, to all transcripts. This is discussed in more detail in the supplementary information and a subsequent publication (Huang and Sanguinetti, 2016). Applying the model did not change the splicing ratios but did reduce variability as transcripts with no reads for a particular junction could now be assigned probabilistically to a spliced/unspliced population (P.Figure S3). This allowed a confidence figure to be assigned to the splicing state of each transcript. Further filtering on confidence was applied to the data set.

82 RPGs and 35 non-RPGs were analysed for their splicing ratio: i.e. probable mRNA FPKM / (probable mRNA FPKM + probable pre-mRNA FPKM) for each time point. These ratios are displayed in P.Figure 5. The bars in the graph are like a histogram but overlapping. In the publication the area under the curve (AUC) of this graph is a simple single number metric, however, it is actually

the AUC (time points 1.5 to 5) divided by the levels at steady state. Noticeable is that RPGs are spliced much more consistently and generally more efficiently than even those non-RPGs that survived the filtering processes. This result was confirmed by RT-qPCR (P.Figure 6B, displayed along with the sequencing profile for these genes, C to E).

4.3.6 Factors Affecting Splicing Kinetics

Transcripts were split into fast and slow splicing groups based on their AUC measure, and features were sought that might contribute to membership of either group (P.Figure 7, S6 and S7). Secondary structure is one of those factors; the more stable the structure the more likely the intron was to be slowly spliced. This was particularly true when only the body of the intron; 5'SS to BPA, is examined. BPA to 3'SS secondary structure has been reported as being beneficial to 3'SS selection in some cases (Plass et al., 2012). The opposite was observed in this study, but only for non-RPGs. Secondary structure could well be beneficial for the less abundant transcripts that did not pass through the filtering process. The length of the intron also influences non-RPGs, albeit minor; faster splicing introns tend to be smaller. The absence of an effect of intron length or BP to 3'SS distance on RPGs could be because there is less variation of these parameters in these genes (this could also be the reason why splice site strengths have little predictive power). Some of these findings were reproduced in follow-on study (using the thiolation procedure in chapter 3), where splice site strength had an effect on splicing fidelity (Aslanzadeh et al., 2018), however, fidelity was not tested in this study.

The effect of secondary structure on splicing can be tested on paralogous RPGs. These genes differ only in their intron sequence and intron secondary structure strength does predict their splicing efficiency (P.Figure 7C).

Poorly spicing RPGs also have an over-abundance of Us in their intron (P.Figure 7D). This is not related to secondary structure, as A shows an opposite effect. Nor is it related to U bias from the thiouracil used to label the

RNA as this does not seem to have affected purification (P.Figure S2 P.Table S2 and discussion in 4.3.1).

Considering all these intron features together (including those discussed in the supplementary material), does have some predictive power for RPGs. The correlation between predicted splicing strength and actual splicing strength is 0.578, an R^2 of 0.334, roughly 33% of the splicing efficiency is explained by the factors chosen. There are factors affecting splicing still to be identified. However, many of the RPGs are probably spliced as well as they can be, the cluster at the right of the scatterplot (reproduced in Figure 1A), are all very efficiently spliced. Spliceosome abundance and maximal splicing rate could be the limit to splicing speed (Hochberg-Laufer et al., 2019) not a factor intrinsic to the transcript. If these are discounted proportionately more points would lie on the slope than presented here, but the small numbers of transcripts reduce the statistical power.

A factor not discussed is whether the branch point A is available to base pair with U2, i.e. masked in a stem or available in a loop. Of the introns for which secondary structure predictions can be made (RNAfold (Lorenz et al., 2011)), 46.4 % of all intronic bases are unpaired, the rest involved in forming a stem. However, in 211 introns the BPA is predicted to be in a loop, whereas 94 are paired in a stem, so 70 % of BPAs being unpaired is remarkable. This has a χ^2 test P value of $<1x10^{-14}$ compared to the average intron pairing rate. Even given the lack of precision of RNA secondary structure prediction, this must be significant. In the fast splicing group, as expected, 71.4 % of BPAs are unpaired, in the slow 64.2 %. However, if just the slowest 10 are examined, only 40 % are unpaired, which approximates to the general rate of pairing in an intron, indicating that these BPAs are not distinguished by secondary structure. This higher BPA pairing proportion has a P value <0.05 (χ^2) compared to the rate of BPA pairing in all introns. These slowest 10 splicing introns form a distinct tail in Figure 1A. Having a BPA hidden in a stem could be an explanation for their slower splicing.

If co-transcriptional splicing occurs, splice sites should be bound by splicing factors and the RNA should not have the freedom to form secondary structure, particularly not long-range interactions involving all the intron seen in YPR187W. For YDR025W (Figure 1C), many of the intronic stems and loops are local and could have space to form secondary structure before a spliceosome component can bind. Secondary structure might only be an important factor if splicing is post-transcriptional (Eperon et al., 1988). However, as most introns are spliced co-transcriptionally, this does not explain why the majority of BPA are not involved in intramolecular base pairing. Coding region secondary structure does affect gene expression and regulation, regions of low Tm RNA being evolutionary conserved (Qi and Frishman, 2017).

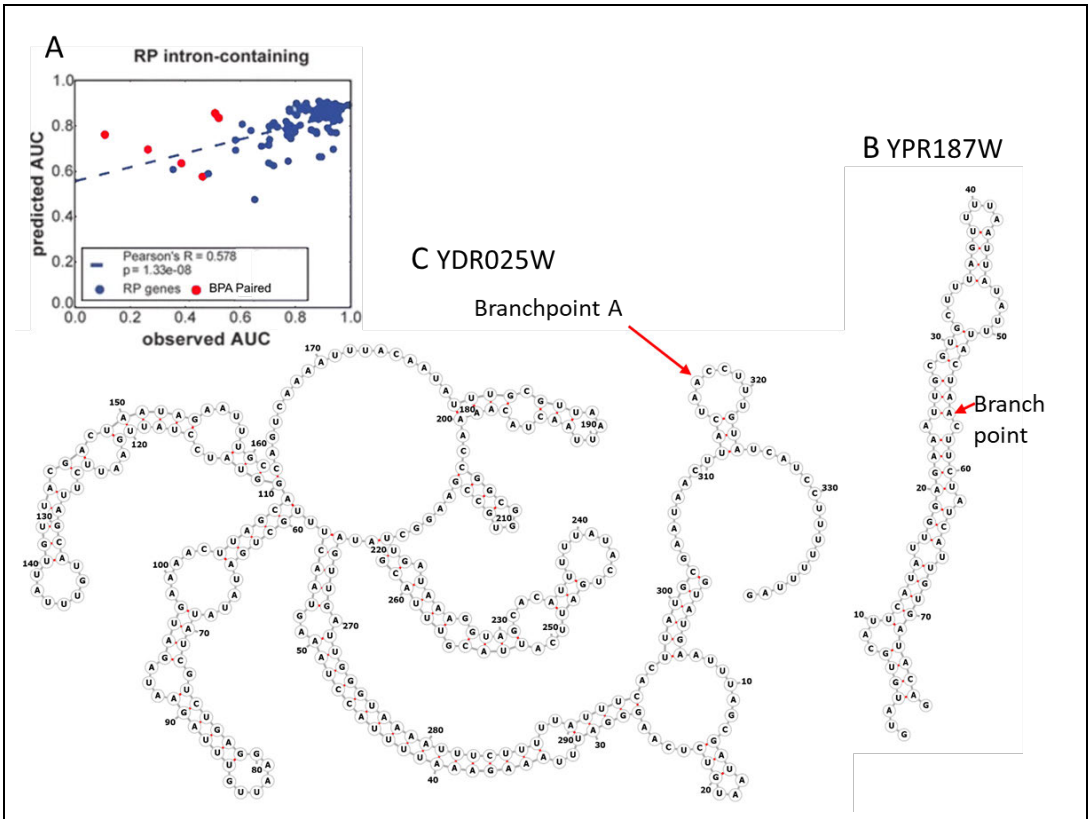


Figure 1: Branchpoint A in a stem structure is deleterious to splicing
A) 6 of the 10 with the lowest observed AUC (indeed 6 of the 8 of the lowest) have a BPA predicted to be base-paired in an intronic stem (red). Only the slowest 10 have had their colour changed to red if paired. From P.Figure 7E
B) Predicted structure (RNAfold) of the intron in YPR187W (*RPO26*), the intron with the lowest AUC, the BPA (indicated) forms part of a stem.
C) Predicted structure of the intron in YDR025W (*RPS11A*), the intron with the lowest AUC but with the BPA in a loop.

4.4 Re-interpretation

The AUC measurement is not the actual area under the curve (but AUC/SS RNA levels), Transcription has an effect: high transcription increasing the AUC. A measure based on the slope would have been better, either the simple gradient, or, as I argued, the point at which the splicing would reach SS levels. However, the AUC measure is easy to calculate, requires no subjective judgements, and in practice produces a useful measure.

4.4.1 Future work

It would be interesting to repeat this experiment with the optimised protocol described in chapter 3, with earlier and more frequent sampling. This would allow much finer resolution and modelling of these processes.

Secondary structure clearly has an important part to play in splice site availability. A future study could examine if freeing a paired BPA, or pairing a free BPA, changes the splicing speed and whether folding round the BS was local or global to the whole intron.

4.4.2 Co-transcriptional Splicing

The data generated in this study were incorporated into the article by Wallace and Beggs (Wallace and Beggs, 2017), examining co-transcriptional splicing. As discussed in this thesis introduction, it combines nascent RNA-seq, NET-seq (Churchman and Weissman, 2012),(Nojima et al., 2015) and SMIT (Carrillo Oesterreich et al., 2016) data with data from this publication. Between these four techniques there is very strong agreement about which pre-mRNAs are co-transcriptionally spliced, particularly for RPGs, and particularly if the 90 % spliced data set is used from the SMIT data.

The introns were categorised into co- versus post-transcriptionally spliced. Most genes were co-transcriptionally processed, only 20 post-transcriptionally spliced. Post-transcriptionally spliced transcripts are likely to be amongst the poorest expressed and therefore not have passed through the filtering in this

chapter's publication, so there may well be more than 20. All the RPGs are predicted to be co-transcriptionally processed and, has been seen here, they seem to be a class apart, even from highly expressed non-RPGs. This has implications for the 6 introns discussed in the section on the accessibility of the BPA. Co-transcriptional splicing implies that the RNA near the BS would be only briefly un-bound by splicing factors, so how secondary structure arises in this region is not clear.

Transcripts with introns known to have splicing regulated by a protein binding the RNA specifically; *RPL30* (Vilardell and Warner, 1997), *RPS14A&B* (Fewell and Woolford, 1999), *YRA1* (Preker and Guthrie, 2006), *DBP2* (Barta and Iggo, 1995) and *RPL9A* (Petibon et al., 2016) are outliers in all assays where they have been detected. For those that can be detected by the experimental systems, degradation mediated by a protein binding to an intron seems to be co-transcriptional also.

YRA1 is interesting in that it shows cell cycle dependant splicing; it encodes a component of the RNA export pathway.

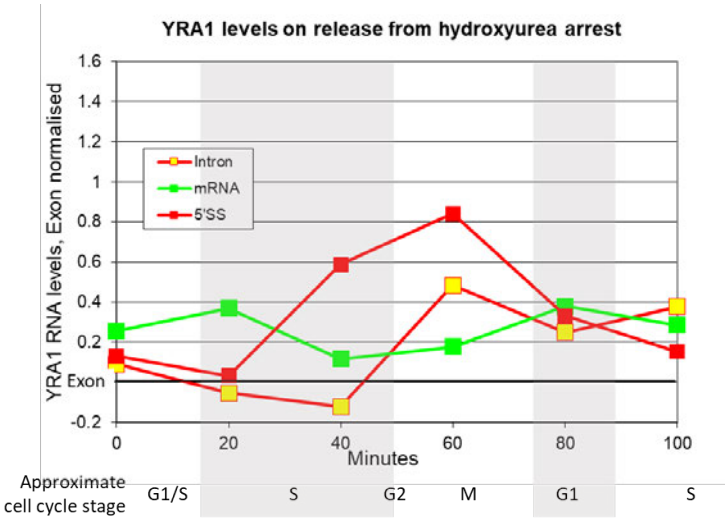


Figure 2: Splicing of YRA1
Microarray determination of *YRA1* spicing on release from hydroxyurea cell cycle arrest. *YRA1* is particularly poorly spliced in G2 and M phase. This result was confirmed by flow cytometry estimation of cell cycle phase and PCR validation of splicing (not shown).

4.4.3 Decay and Export

This technique was used by another group (with which I collaborated), to examine degradation when export is blocked (Tudek et al., 2018). Export was blocked by the rapid depletion of Mex67p, a component of the export heterodimer with Mtr2p. Yra1p is an adaptor protein, as is the nuclear poly A binding protein Nab2p. Depletion of Mex67p resulted in rapid nuclear accumulation of polyadenylated (pA+) mRNA, followed by loss of these transcripts and ultimately cell death. RNA fate was assayed by 4tU using the protocol described in this chapter and showed that degradation of pA+ RNA was not the cause of the loss, but rather decreased production of pA+ RNAs. That is, the transcripts were made but not polyadenylated and therefore degraded rapidly. Correct termination at the 3' end was disrupted for a few genes leading to run on transcription, suggesting a link between transcription and export. Similar results were obtained when other export factor depletions were made.

Nab2p over-expression partially reverses this pA+ RNA loss. The model proposed is that nuclear export block leads to pA+ RNA becoming trapped in the nucleus, Nab2p is sequestered on these poly(A) tails so none is free to bind all the new transcripts. Therefore poly(A) tails are not being added to the new transcripts, possibly because CFI/CPF is not displaced from the 3' ends by Nab2p, this leads to occasional run on transcription and degradation of unpolyadenylated RNAs

4.5 Contribution

The labelling protocol was developed by myself, see chapter 3. Much of the cell growth and sampling was performed collaboratively between myself and Jane Reid, a PhD student I was supervising. nsRNA preparation was by Jane under my tutelage and the RNA Sequencing was done by Jane. Analysis was collaborative, between myself, Jane and Yuanhua Huang, Yuanhua did the mathematical bioinformatics. I performed all the RT-qPCRs.

All additional data presented in this chapter are my own, as is any re-interpretation of the results and conclusions from the publication.

4.6 References

Most references are in the publication, a few additional ones are listed here.

Aslanzadeh, V., Huang, Y., Sanguinetti, G., and Beggs, J.D. (2018). Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. Genome Res. *28*, 203–213.

Barrass, J.D., and Beggs, J.D. (2019). Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE J. Vis. Exp. e59952.

Barta, I., and Iggo, R. (1995). Autoregulation of expression of the yeast Dbp2p "DEAD-box" protein is mediated by sequences in the conserved DBP2 intron. EMBO J *14*, 3800–3808.

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165* 372-381.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism11Edited by J. Karn. J. Mol. Biol. *284*, 975–988.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al *Chapter 4*, Unit 4.14.1-17.

Eperon, L.P., Graham, I.R., Griffiths, A.D., and Eperon, I.C. (1988). Effects of RNA secondary structure on alternative splicing of Pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? Cell *54*, 393–401.

Fewell, S.W., and Woolford, J.L. (1999). Ribosomal protein S14 of Saccharomyces cerevisiae regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. Mol. Cell. Biol. *19*, 826–834.

Grzechnik, P., Szczepaniak, S.A., Dhir, S., Pastucha, A., Parslow, H., Matuszek, Z., Mischo, H.E., Kufel, J., and Proudfoot, N.J. (2018). Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. Nat. Commun. *9*, 1783.

Hochberg-Laufer, H., Neufeld, N., Brody, Y., Nadav-Eliyahu, S., Ben-Yishay, R., and Shav-Tal, Y. (2019). Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. PLOS Genet. *15*, e1008459.

Huang, Y., and Sanguinetti, G. (2016). Statistical modeling of isoform splicing dynamics from RNA-seq time series data. Bioinformatics *32*, 2965–2972.

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

Petibon, C., Parenteau, J., Catala, M., and Elela, S.A. (2016). Introns regulate the production of ribosomal proteins by modulating splicing of duplicated ribosomal protein genes. Nucleic Acids Res. *44*, 3878–3891.

Plass, M., Codony-Servat, C., Ferreira, P.G., Vilardell, J., and Eyras, E. (2012). RNA secondary structure mediates alternative 3′ss selection in Saccharomyces cerevisiae. RNA *18*, 1103–1115.

Preker, P.J., and Guthrie, C. (2006). Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. RNA *12*, 994–1006.

Qi, F., and Frishman, D. (2017). Melting temperature highlights functionally important RNA structure and sequence elements in yeast mRNA coding regions. Nucleic Acids Res. *45*, 6109–6118.

Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., and Jensen, T.H. (2018). A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA. Cell Rep. *24*, 2457-2467.e7.

Vilardell, J., and Warner, J.R. (1997). Ribosomal protein L32 of Saccharomyces cerevisiae influences both the splicing of its own transcript and the processing of rRNA. Mol Cell Bio *17*, 1959–1965.

Wallace, E.W.J., and Beggs, J.D. (2017). Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. RNA *23*, 601–610.

4.7 Reprint

Barrass *et al. Genome Biology* (2015) 16:282
DOI 10.1186/s13059-015-0848-1

Genome Biology

RESEARCH Open Access

Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling

J. David Barrass[1†], Jane E. A. Reid[1†], Yuanhua Huang[2†], Ralph D. Hector[3,1], Guido Sanguinetti[2,3], Jean D. Beggs[1*] and Sander Granneman[3*]

Abstract

Background: RNA levels detected at steady state are the consequence of multiple dynamic processes within the cell. In addition to synthesis and decay, transcripts undergo processing. Metabolic tagging with a nucleotide analog is one way of determining the relative contributions of synthesis, decay and conversion processes globally.

Results: By improving 4-thiouracil labeling of RNA in *Saccharomyces cerevisiae* we were able to isolate RNA produced during as little as 1 minute, allowing the detection of nascent pervasive transcription. Nascent RNA labeled for 1.5, 2.5 or 5 minutes was isolated and analyzed by reverse transcriptase-quantitative polymerase chain reaction and RNA sequencing. High kinetic resolution enabled detection and analysis of short-lived non-coding RNAs as well as intron-containing pre-mRNAs in wild-type yeast. From these data we measured the relative stability of pre-mRNA species with different high turnover rates and investigated potential correlations with sequence features.

Conclusions: Our analysis of non-coding RNAs reveals a highly significant association between non-coding RNA stability, transcript length and predicted secondary structure. Our quantitative analysis of the kinetics of pre-mRNA splicing in yeast reveals that ribosomal protein transcripts are more efficiently spliced if they contain intron secondary structures that are predicted to be less stable. These data, in combination with previous results, indicate that there is an optimal range of stability of intron secondary structures that allows for rapid splicing.

Keywords: Cryptic transcripts, Metabolic labeling, RNA sequencing, Splicing, Yeast

Background

The RNA levels detected in cells at steady state are the consequence of multiple dynamic processes within the cell. Eukaryotic genomes are pervasively transcribed, but the accumulation of many transcripts is limited by processes that regulate their synthesis and decay [1]. In addition, most primary transcripts undergo processing events. For example, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), transfer RNAs (tRNAs), microRNAs (miRNAs), small interfering RNAs (siRNAs) and some long non-coding RNAs (lncRNAs) are often produced by the post-transcriptional processing of short-lived longer transcripts that are more readily detected in the absence of degradation or processing factors [2]. A number of different classes of lncRNAs have been described in eukaryotes [3, 4]. The best characterized examples in yeast are the stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs) and Xrn1-sensitive unstable transcripts (XUTs), three types of lncRNAs thought to have different stabilities [4]. SUTs are readily detectable in wild-type cells and share many similarities with messenger RNAs (mRNAs), whereas CUTs are generally more unstable and frequently only detectable in cells lacking the nuclear exosome components. XUTs are likely (primarily) degraded

* Correspondence: jbeggs@ed.ac.uk; sgrannem@staffmail.ed.ac.uk
†Equal contributors
[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK
[3]Centre for Synthetic and Systems Biology (SynthSys), University of Edinburgh, Edinburgh EH9 3BF, UK
Full list of author information is available at the end of the article

in the cytoplasm because they accumulate in the absence of the cytoplasmic exoribonuclease Xrn1 [5].

In the case of intron-containing genes, the level of mature transcripts is influenced by splicing, as well as by synthesis and decay. Splicing of pre-mRNAs (precursors of messenger RNAs) [6] occurs in the nucleus, often co-transcriptionally [7]. The spliced mRNA is exported to the cytoplasm where it can be translated, whereas the excised intron, which has a branched, lariat structure, is rapidly debranched and degraded. Measurements of in vivo RNA processing rates and efficiencies depend on the ability to estimate the levels of the unprocessed precursors and processing intermediates in cell extracts; however, this is challenging because they are highly transient and present in low abundance in wild-type cells at steady state.

A diverse range of methods has been used to measure the kinetics of RNA processing. One approach involves inhibiting RNA polymerase II (Pol II) activity with a transcription inhibitor such as 5,6-dichloro-1-β-D-ribobenzimidazole, then removing this inhibition to promote synchronized transcription. For example, measuring the delay after Pol II reaches the 3′ ends of introns before the spliced flanking exons can be detected provides an estimate of the time for splicing [8]. Splicing kinetics have also been estimated in human cells using fluorescence recovery after photobleaching and live cell imaging, by measuring either the amount of time that fluorescently tagged spliceosomal small nuclear ribonucleoproteins associate with transcripts, or the time taken for an intron-tethered green fluorescent protein-fusion protein to be removed by splicing [9]. By these diverse approaches, estimates of the time taken to splice introns in human cells have varied between 30 s and 5–10 min. For budding yeast, Alexander et al. [10] used an inducible reporter gene and high resolution kinetic analysis of RNA by quantitative reverse transcription polymerase chain reaction (RT-qPCR), estimating that splicing occurred within 30s of the production of the pre-mRNA. There are also computational approaches for predicting the speed of RNA processing events [11].

Metabolic labeling is a way of proportionally enriching classes of RNA that are difficult to detect in wild-type cells at steady state. For example, 4-thiouridine (4sU) has been used to label newly synthesized transcripts in human cells [12] and budding yeast cells [13] followed by biotinylation of the thiolated RNA to allow its selective recovery for microarray analysis. Neymotin et al. [14] used 4-thiouracil (4tU) labeling of RNA in budding yeast followed by RNA sequencing o determine RNA degradation rates, from which synthesis rates were derived.

Labeling with 4sU has been used to measure RNA synthesis, decay and splicing rates in human and yeast; however, the shortest labeling time was 3 min, by which time a substantial fraction of the newly transcribed RNA was already spliced or degraded [13, 15]. Therefore, to be able to measure RNA processing rates with higher accuracy and resolution transcriptome-wide, we have developed an extremely short (as little as 60 s) 4tU RNA labeling protocol and combined it with high-throughput RNA sequencing (RNA-seq). We demonstrate that our method (4tU-seq) readily detects low abundance and labile transcripts in wild-type cells that are normally detected only in cells that are defective in RNA degradation. Our data show that at such short times, lncRNA degradation kinetics depart significantly from first-order, and quantitatively associate lncRNA turnover with structural features of the transcripts. Also, using 4tU-seq we could, for the first time, measure relative pre-mRNA splicing kinetics transcriptome-wide in budding yeast. Unexpectedly, our results show that fast splicing of intron-containing ribosomal protein mRNAs largely depends on the degree of secondary structure between the 5′ splice site (5′ss) and branch point (BP) sequence and indicate that there is an optimal range of stability of intron secondary structures and base composition that allows for rapid splicing.

Results and discussion

Thiolated RNA can be efficiently recovered after very brief 4tU labeling

To isolate short-lived RNA species from the yeast *Saccharomyces cerevisiae*, we incubated cells with 4tU for very short periods, extracted the RNA, and treated it with a thio-reactive reagent to biotinylate the newly synthesized transcripts that contain thiol moieties [12, 16]. The biotinylated RNA was then affinity-purified with streptavidin. To improve 4tU incorporation during very short periods, the uptake of 4tU by yeast cells was enhanced by overexpressing the *FUI1* permease gene from a plasmid [17, 18]. Cell metabolism was rapidly halted by snap-freezing the labeled cells directly in very cold methanol, which is crucial for the recovery of short-lived RNAs [10]. Furthermore, each stage of the RNA isolation was carefully optimized to reduce background and maximize yield, in particular by using a modified binding and wash buffer (detailed in "Methods"). We used 4tU rather than 4sU for our studies because 4tU is much less expensive and gives very comparable incorporation rates (data not shown).

There was a linear increase in the yield of thiolated RNA over short labeling times up to 5 min, after which a component of the system became limiting (Fig. 1a; $R^2 = 0.99$). Moreover, labeling for only 1.5 min was sufficient to achieve at least 2-fold enrichment over the background (yield from an unlabeled sample). By fitting a line to the data to indicate background levels, we

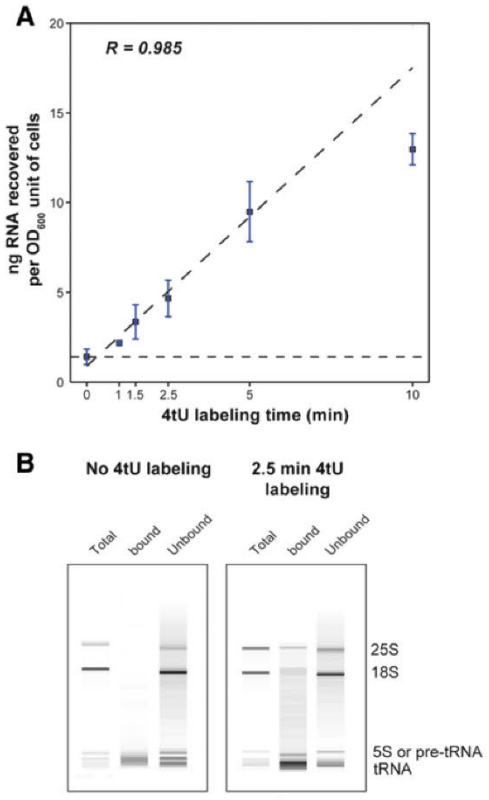


Fig. 1 RNA yield increases with labeling time. **a** Plot of total yield of RNA recovered in nanograms per OD_{600} unit of cells against labeling time in minutes. During the first five minutes of labeling, yield increases linearly with labeling time ($R^2 = 0.985$). Some RNA is recovered when 4-thiouracil (*4tU*) is not added to the culture (0 min). This is essentially background RNA that non-specifically bound to the magnetic beads during the isolation procedure. The *horizontal dashed line* indicates the level of background for this experiment. The actual amounts of RNA recovered from 600 ml culture (average of two experiments; $R^2 = 0.960$) were: 0.50 μg (0 min), 0.77 μg (1 min), 1.16 μg (1.5 min), 3.30 μg (5 min) and 4.52 μg (10 min). **b** Agilent Bioanalyzer trace demonstrating the qualitative differences between the different fractions produced during the isolation of newly synthesized RNA. The *left panel* displays the results when no 4tU is added while the *right panel* shows 2.5 min labeling. The nonspecific background can be seen to be different from the total RNA comprising mostly short fragments. *tRNA* transfer RNA. 25S, 18S and 5S indicate mature ribosomal RNA species.

deduced that the estimated time required before any 4tU was incorporated was about 30 s (Fig. 1a). Bioanalyzer analysis of mock samples indicated that most of the background consisted of short RNA species (Fig. 1b), which were mostly highly abundant tRNAs (Additional file 1: Figure S1).

4tU-seq proportionally enriches low stability species of RNA

We next performed RNA-seq on thiolated RNA (4tU-seq) isolated after 1.5, 2.5 and 5 min of 4tU labeling, on unlabeled control samples (i.e., "background") and on rRNA-depleted total RNA, with all experiments performed at least twice. Additional file 1: Tables S1 to S14 provide all the results of our 4tU-seq data bioinformatics analyses. Additional file 1: Table S1 lists the total number of uniquely mapped reads for each sample. Notably, for the majority of RNA species we did not observe a significant correlation between the fraction of uridines in the transcript and the read coverage or RNA half-life (Additional file 1: Table S2 and Figure S2). However, this was not the case for snRNAs, which had a small sample size (only six) and are renowned for being U-rich (Additional file 1: Table S2).

We then used DESeq2 [19] to calculate the enrichment of different classes of RNA in the 4tU-labeled samples relative to the total RNA samples (see "Methods" for more details on the differential analyses of transcript abundance). In the 1.5-min 4tU-seq samples, a high proportion (37 %) of intron-containing transcripts were significantly enriched (adjusted $p < 0.05$; Fig. 2a), with substantially less enrichment seen in the 5-min samples. This is not surprising given that longer labeling times approach the steady state situation, where more of the labeled transcripts are spliced, such that the proportion of intron-containing transcripts is reduced. This illustrates the benefit of labeling for extremely short times. To our knowledge, 90 s is the shortest labeling period after which RNA-seq has been performed [14, 20].

Notably, many CUTs, SUTs and XUTs were significantly enriched in the 4tU samples relative to total RNA, with the degradation of these species appearing relatively slow compared to pre-mRNA splicing (discussed below).

Many non-coding RNAs (ncRNAs) are matured by cleavage and/or trimming of precursors at the 5′ and/or 3′ ends and the precursors were very well represented in our 4tU-seq data. Pre-ribosomal RNAs (pre-rRNAs) were readily detectable after 1.5 min of labeling, as judged by the accumulation of reads mapping to the external-transcribed and internal-transcribed spacer of the rDNA unit (Fig. 2b). The processing of these spacer regions seems to be slow given that they were still abundantly detectable after 5 min of labeling. It has been reported that incubation of human cells with concentrations of 4sU ≥50 μM over long periods of time (>1 h) causes a nucleolar stress response and inhibition of the production and processing of rRNA [21]. It is not known whether 4tU incubation for the very short times used here affects rRNA processing and, because rRNA was depleted from our total RNA samples before sequencing, we did not calculate relative pre-rRNA processing speeds.

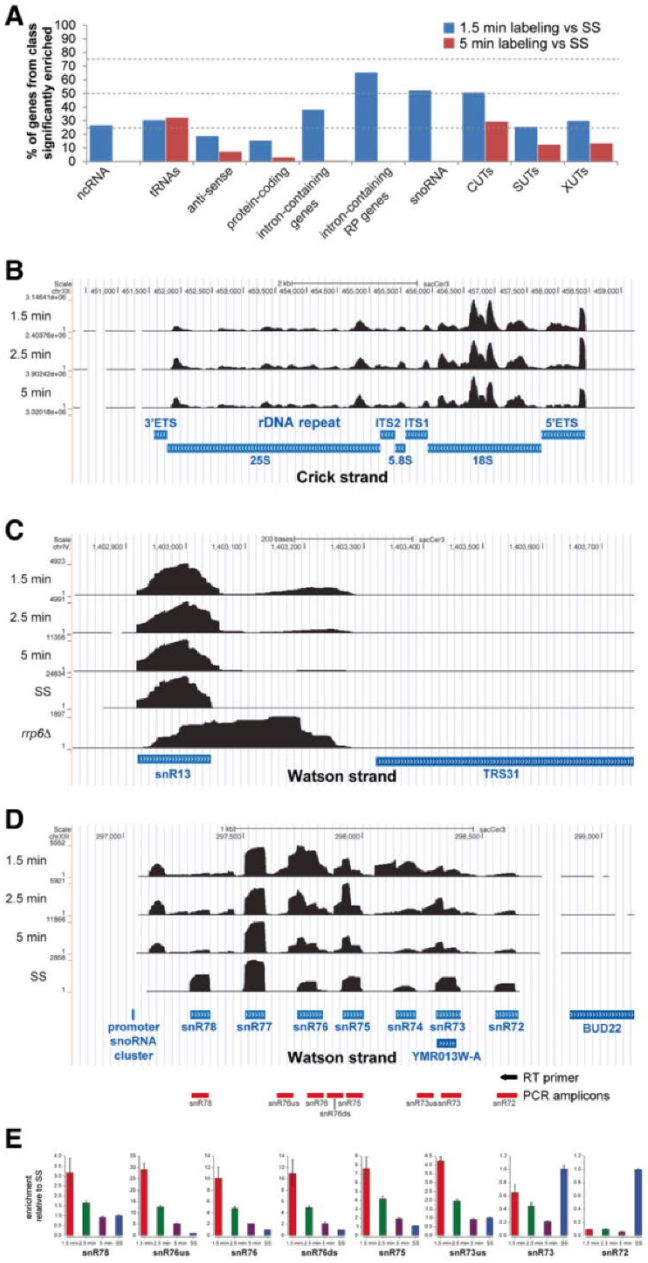


Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Short labeling times proportionally enrich unstable transcripts. **a** DESeq2 [19] was used to identify features significantly enriched in 4tU-seq data from short labeling times (1.5 min and 5 min) compared to total RNA. The figure displays the percentage of transcripts in each category that was found to be significantly enriched (DESeq2 adjusted $p < 0.05$). For the DESeq analyses, all reads were considered. Thus for the intron-containing mRNAs we used reads that mapped to both introns and exons. **b-d** UCSC genome browser screen shots showing the change in distribution of reads at different labeling times (Y-axis), with annotation below in *blue*. *SS* indicates steady-state levels, generated by sequencing total RNA. **b** 4tU detects pre-rRNA precursors. Note that the total RNA sample is not shown because it was rRNA depleted. **c** 4tU-seq detects 3′ extended snR13 species. Data from an rrp6Δ strain are displayed for qualitative comparison. **d** Polycistronic precursor from which multiple snoRNAs are processed. Blue boxes represent the annotated mature snoRNAs. **e** Real time (*RT*) quantitative polymerase chain reaction (*PCR*) validation of the 4tU-seq results shown in (**d**). For the RT reaction, a reverse transcriptase primer was used that was complementary to the 3′ end of the snR72 snoRNA. This cDNA was then used to amplify the different amplicons shown below each bar plot (see the illustration in (**d**) for what each amplicon represents). The data were then normalized to the results obtained with rRNA-depleted total RNA (SS). *5′ETS 5′ external transcribed spacer*, *3′ ETS 3′ external transcribed spacer*, *4tU* 4-thiouracil, *CUTs* cryptic unstable transcripts, *ITS internal transcribed spacer*, *ncRNA* non-coding RNA, *RP* ribosomal protein, *snRNA* small nuclear RNA, *snoRNA* small nucleolar RNA, *SUTs* stable unannotated transcripts, *tRNA* transfer RNA

The snR13 snoRNA is processed from a 3′ extended precursor [22] that accumulates in nuclear RNA surveillance mutants [23, 24]. Our 4tU-seq data suggest that most 3′ extended snR13 species are processed within 5 min of 4tU labeling. Another striking example from our 4tU-seq data was the rapid processing of snoRNAs from the snR72-snR78 polycistronic transcript [25] (Fig. 2d), which we confirmed by reverse transcriptase quantitative PCR (RT-qPCR; Fig. 2e). These results clearly demonstrate the potential of very short 4tU-labeling experiments for measuring the processing kinetics of short-lived RNA species.

Degradation rates of cryptic transcripts correlates with transcript length and secondary structure

A comparison of the enrichment of the different classes of cryptic transcripts for the 1.5 and 5 min labeling times relative to the total RNA samples indicated that CUTs, SUTs and XUTs were all readily detected and decayed more slowly than intron-containing transcripts and snoRNAs (Fig. 2a, Table 1, Additional file 1: Figure S1). The heat maps in Fig. 3a, b show hierarchically clustered normalized read data for 887 CUTs and for 823 SUTs [26] at 1.5 min, 2.5 min, 5.0 min and steady state. They clearly show the slow decay of both types of transcripts. The expression levels of CUTs and SUTs were generally similar at early labeling times but significantly different at steady states ($p < 0.05$; Fig. 3c). Interestingly, the levels of both CUTs and SUTs rose rapidly and remained approximately constant for the first 5 min at a level that was considerably higher than the steady state (Figs 2a and 3c, Additional file 1: Figure S1). This seems incompatible with the widely adopted first order kinetic assumption [27] and suggests that processing of these RNA species involves multiple steps, inducing delays that are comparable with the sampling times used in this study.

Table 1 Comparison of degradation dynamics between 925 cryptic unstable transcripts (*CUTs*) and 847 stable unannotated transcripts (*SUTs*)

Time gaps	CUT decrease	SUT decrease	*p*-value (t test)
1.5–2.5 min	0.31	0.32	0.79
1.5–5.0 min	0.63	0.68	0.84
1.5 min steady state	3.41	2.86	3.9×10^{-10}

The first column shows the gap between the two time points considered. The second column is the median decrease between the two time points for CUTs [computed as log2(fragments per kilobase per million reads)]. The third column shows the same quantity for SUTs. The fourth is the t-test *p*-values of the differences between CUTs and SUTs decreases.

The 4tU-seq data for three example CUTs (Fig. 3d) were compared with RT-qPCR analysis of the same three 4tU-labeled CUTs (Fig. 3e), and showed good agreement between the two methods. Thus, despite the fact that these RNA species are labeled as "unstable," processing of many CUTs appears to be relatively slow.

We then sought to quantitatively explain the observed processing kinetics of CUTs and SUTs from sequence and structural features. As a measure of decay rate for each transcript, we considered the log ratio of the average expression level at the early times (1.5, 2.5 and 5 min) over the steady state expression level (Fig. 4a). The rationale behind this choice was the following: given the strong deviation from first order kinetics at early times, we assumed that early measurements were effectively proportional to RNA production rate, whereas the steady state expression levels were approximately equal to the ratio of production rate to decay rate (assuming that the initial deviation from first order kinetics becomes unimportant over long times). We observed that this decay rate was significantly negatively associated with transcript length and predicted secondary structure (Fig. 4b); in other words, long and/or highly structured transcripts were more stable. This is consistent with the model that many ncRNAs are subject to early termination by the Nrd1-dependent pathway and rapid degradation by the nuclear exosome [1–3, 4, 28] and/or are degraded in the cytoplasm (e.g., by Xrn1 [5]). Because SUTs are significantly longer than CUTs (Fig. 4c), our data provide a possible explanation for why SUTs are generally more stable than CUTs. We note that, in a regime where transcription is sporadic, transcript length could be expected to be positively associated with

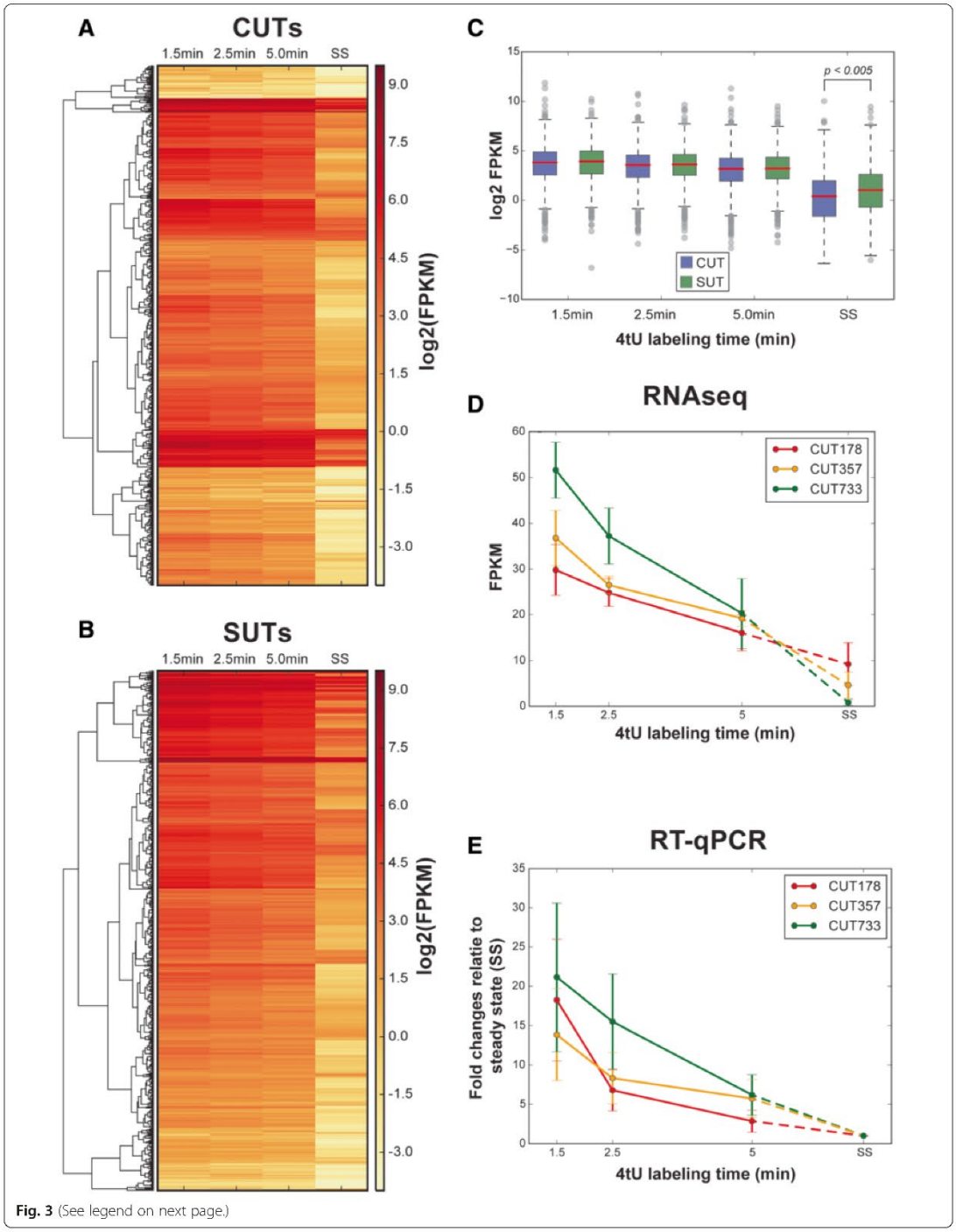


Fig. 3 (See legend on next page.)

(See figure on previous page.)
Fig. 3 Cryptic unstable transcripts (*CUTs*) and stable unannotated transcripts (*SUTs*). **a** Heat map of fragments per kilobase per million reads (*FPKM*) at log2 scale for 887 CUTs at 1.5 min, 2.5 min, 5.0 min and steady state (*SS*). The hierarchical clustering is based on complete similarity between the FPKMs of two CUTs at the four time points. **b** The same heat map for 823 SUTs. **c** Comparison between log2 FPKM of 887 CUTs and 823 SUTs at 1.5 min, 2.5 min, 5.0 min and steady state. The levels of CUTs and SUTs are similar at initial time points but significantly different at steady states. **d** FPKM changes for three example CUTs measured by RNA-seq. **e** Levels of the same three CUTs relative to steady state, measured by reverse transcription quantitative polymerase chain reaction (*RT-qPCR*). *4tU* 4-thiouracil

recovery at early time points, because longer transcription times would increase the probability of a transcript being labeled. Our analyses failed to reveal a significant effect of transcript length on the recovery of relatively short transcripts such as CUTs and SUTs at all time points (Additional file 1: Figure S3). Indeed, a bias toward longer transcripts at earlier labeling times would lead to a negative association of length with stability (longer transcripts would appear to have higher early expression and hence higher decay rate), which is the opposite of what we observed.

When attempting to fit a first order kinetic model to the data, the resulting degradation rates were uncorrelated to transcript length or structural features (data not shown), further underlying the inappropriateness of the first order assumption at these extremely short labeling times. Furthermore, using predicted secondary structure and transcript length alone, we were able to train a machine learning classifier to classify CUTs and SUTs with reasonable accuracy (Fig. 4d; area under receiver operating characteristic curve was 0.68 for CUTs and 0.73 for SUTs, random baseline 0.5).

Very short 4tU labeling enables accurate measurements of splicing kinetics

In the case of intron-containing transcripts, the extent of RNA splicing can be determined by several alternative approaches [7], for example, from sequence reads that span intron–exon boundaries (splice sites) relative to reads that cross splice exon junctions [29], or from the number of intron reads relative to total gene reads [20]. Importantly, the transcription rate for any given gene should be constant during short periods of growth at steady state and will affect intron and exon production similarly, allowing relative splicing speeds to be determined by comparing pre-mRNA and exon levels at different labeling times without the influence of transcription rate. In vivo, spliceosome assembly occurs largely co-transcriptionally [7] and, theoretically, splicing catalysis could occur on a nascent transcript as soon as the 3′ss exits the polymerase. Additionally, because the 4tU-label can be incorporated at any position in the transcript at the time when the label is added, spliced mRNAs can be labeled even at the earliest time point. This could explain the very rapid splicing of some transcripts. Splicing also occurs in competition with pre-mRNA degradation [30, 31]; however, because degradation removes both the intron and the exon, this would not affect our estimates of splicing ratios.

In this work, we used a probabilistic model to estimate splicing ratios, that is, the proportion of spliced mRNA out of the total RNA for each transcript. This method was a modification of the MISO model [32] for quantification of splicing isoforms. Briefly, the method used information from all mapped reads by introducing a latent categorical variable for each read: the identity (i.e., whether it came from mature or precursor mRNA). The method then used computational statistical tools to compute a posterior probability of the identity of each read, therefore providing a quantification of mature and precursor mRNA. It is important to observe that in some cases (e.g., boundary or junction reads) the identity of a read is unambiguous, and indeed many direct methods only use unambiguously assigned reads; however, it was convincingly shown [32] that using information from all reads leads to more accurate estimations of isoform abundance. For more details, and for a comparison of different estimation methods, see Additional file 1: Data and Methods.

A major benefit of using a probabilistic model for estimating the abundance of precursor and mature mRNA lies in the possibility of obtaining posterior confidence intervals (CI) on the splicing ratios, which can then be used to filter noise in a principled way. In the analysis, we only retained genes with 95 % CI < 0.3, filtering out genes for which reliable estimation was not possible (primarily due to low sequence coverage). This filtering improved the correlation between replicates from 0.757 to 0.864 (see Additional file 1: Tables S3, S4 and S7). We also used a simulation to compare quantitatively the performance of the probabilistic model against the two direct methods that use the splice junction and intron boundary reads only, or intron reads and exon reads. The simulation showed that the probabilistic model yielded considerably lower variance results than the direct methods (Additional file 1: Figure S4), and therefore likely generated better results, especially when analyzing genes with low read coverage.

In this 4tU-seq analysis, 187 intron-containing transcripts were selected that had a fragments per kilobase per million reads score (FPKM) of >10, that did not encode snoRNAs in the introns, and that only contained a single intron (to simplify the data analyses). Following

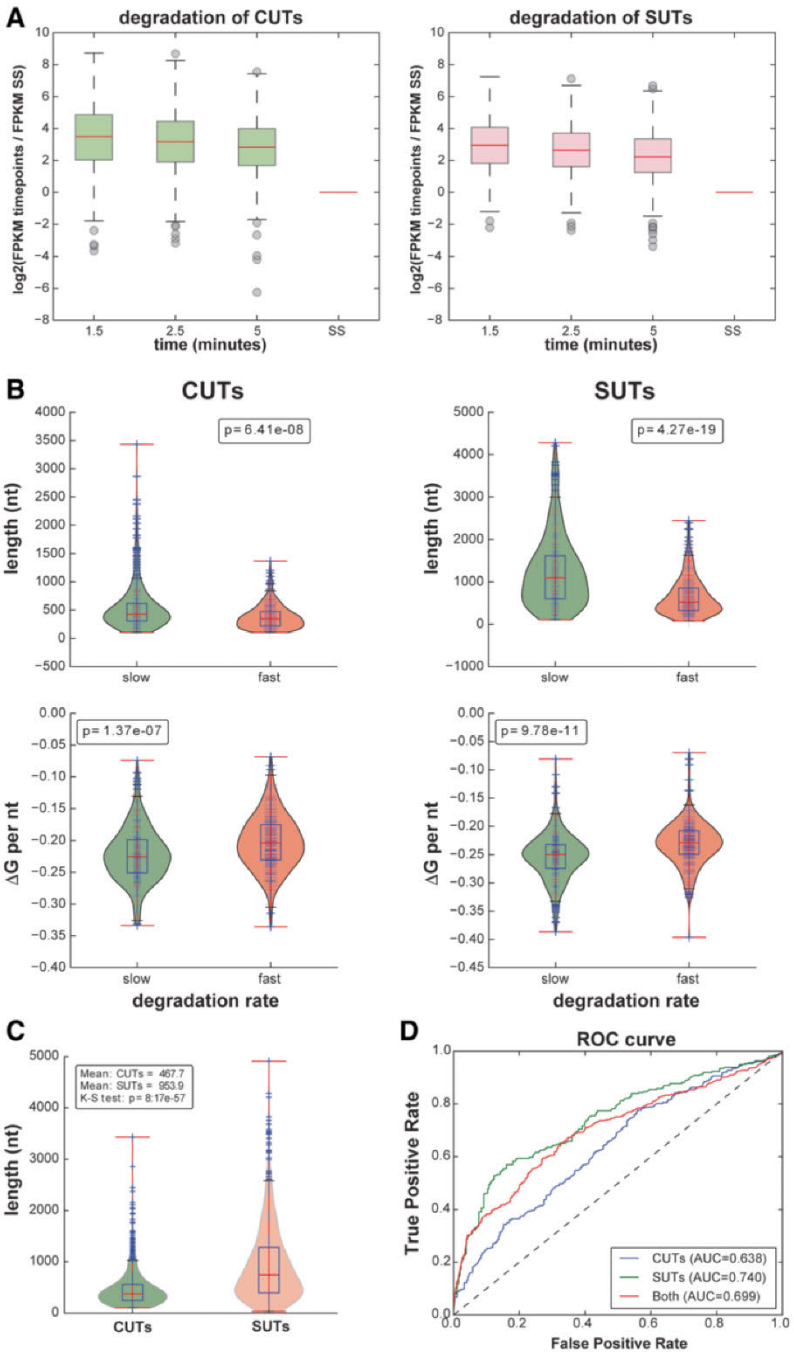


Fig. 4 (See legend on next page.)

(See figure on previous page.)
Fig. 4 Analysis of degradation of cryptic unstable transcripts (*CUTs*) and stable unannotated transcripts (*SUTs*). **a** The log2 scaled ratio of CUTs fragments per kilobase per million reads (*FPKM*) normalized to the steady state (*SS*) levels. These ratios were used to quantify the degradation rate. A weighted average of the three nascent ratios was used to quantify the degradation rate, that is, the higher the ratio at the nascent points, the faster the degradation. **b** Comparison of features between the fastest-degrading third and slowest-degrading third of CUTs/SUTs. Features include average secondary structure free folding energy for each nucleotide, and transcript length. **c** SUTs are significantly longer than CUTs. Comparison of CUT and SUT transcript length distribution using the Kolmogorov–Smirnov (*K–S*) test. **d** The binary classification between the fastest-degrading third of transcripts and the slowest-degrading third of transcripts using ΔG per nucleotide (*nt*), transcript length, ΔG of ±15 nt around the start site and ΔG of ±15 nt around the stop site. The receiver operating characteristic (*ROC*) curves shows the data for the CUTs, SUTs or both with 10-folder cross-validation via a naive Bayes classifier. The area under the curve (*AUC*) is used to represent the prediction performance

posterior CI filtering, data for 82 ribosomal protein (RP) and 35 non-RP intron-containing genes were retained for the splicing ratio analysis (Additional file 1: Table S7). Figure 5 shows the mean splicing ratio of the three replicates at different time points for the 35 non-RP and 82 RP intron-containing genes, in which transcripts are ranked by speed of splicing from fastest (top) to slowest (bottom).

We also estimated the amount of background unlabeled (pre-)mRNA in our samples to determine to what extent this could influence our estimated splicing speeds. For these calculations we used RNA sequencing data generated from "background" or "0" samples (0 time point in Fig. 1a), which were derived from RNA that non-specifically bound to the magnetic beads during the isolation procedure (see "Methods" for details). We considered the intronic and exonic FPKM of the 250 intron-containing protein-coding genes in the background and 1.5-min samples. The middle column in Additional file 1: Figure S1 shows that the intronic FPKM percentages were 7.5 % and 46.1 % in the background and 1.5-min samples, respectively. Intuitively, the presence of background RNA will tilt this balance towards spliced mRNA, because unlabeled RNA is overwhelmingly exonic in the background samples (Additional file 1: Figure S1). Using these data, we estimated the background proportion using the difference between the 4tU-labeled intronic RNA fraction and the actually measured fraction. Theoretically, the intronic FPKM percentage in the signal should be lower than 50 %, with this bound being attained if spliced intronic RNA decayed at the same rate as mature mRNA (clearly a worst case scenario). Using this theoretical assumption, we estimated the upper bound on the background mRNA levels in the 1.5-min samples to be <9.1 % (see details in "Methods"). We conclude that this low level of mRNA background may have very slightly affected our kinetic estimates of splicing speeds but will not have affected the rankings (Fig. 5).

To validate some of our 4tU data, 4tU-RT-qPCR was performed on transcripts of three well-expressed ribosomal protein genes, *RPS13*, *RPL28* and *RPL39* (Fig. 6a, b). The levels of the 5′ss boundary amplicons (corresponding to unspliced pre-mRNAs) were normalized to the values for the corresponding exon 2 amplicon (Fig. 6a). The resulting values provided a measure of the proportion that corresponded to pre-mRNA. As the labeling time increases, the amount of exon 2 should remain more or less constant but the proportion of precursor should change. After 1.5 min of 4tU labeling, the proportion of pre-mRNA was high and decreased with labeling time, gradually approaching the level at steady state. The time required to reach steady state was transcript dependent, being approximately 5 min for RPL39, which was spliced relatively quickly, and longer for the others, in agreement with the rank order in Fig. 5b. These results confirm the enhanced recovery of unspliced pre-mRNAs by the 4tU-seq approach and its ability to detect different rates of pre-mRNA splicing for different transcripts. The results also validate the probabilistic model used.

One drawback of using intron read counts to measure splicing rates is that it is not always possible to distinguish between reads originating from pre-mRNA introns or intron lariats. This could potentially lead to an underestimation of splicing rates for some mRNA transcripts. However, based on the following results, we do not think that lariat reads significantly contributed to the splicing speeds reported here: several attempts were made to identify lariats in the 4tU-seq data using existing strategies [33] but only a handful of lariats could be detected. This is not surprising, because sequencing lariats tends to require use of *dbr1* mutants that lack lariat debranching activity and dedicated library preparation [33, 34]. Furthermore, we found a very high correlation between intronic read counts and counts for reads that overlap 5′ ss–intron boundaries (Additional file 1: Figure S5). Therefore, we conclude that our use of intron versus exon reads accurately measures splicing rates.

A/U richness and secondary structure of the intron affects ribosomal protein splicing kinetics

Our results show that different introns were spliced at different rates based on area under the curve (AUC) calculations (see "Methods," and scores in Fig. 5). Various features of introns could impact their speed of splicing. These include how close to consensus are the 5′ss, 3′ss and BP sequences as well as the strength of the

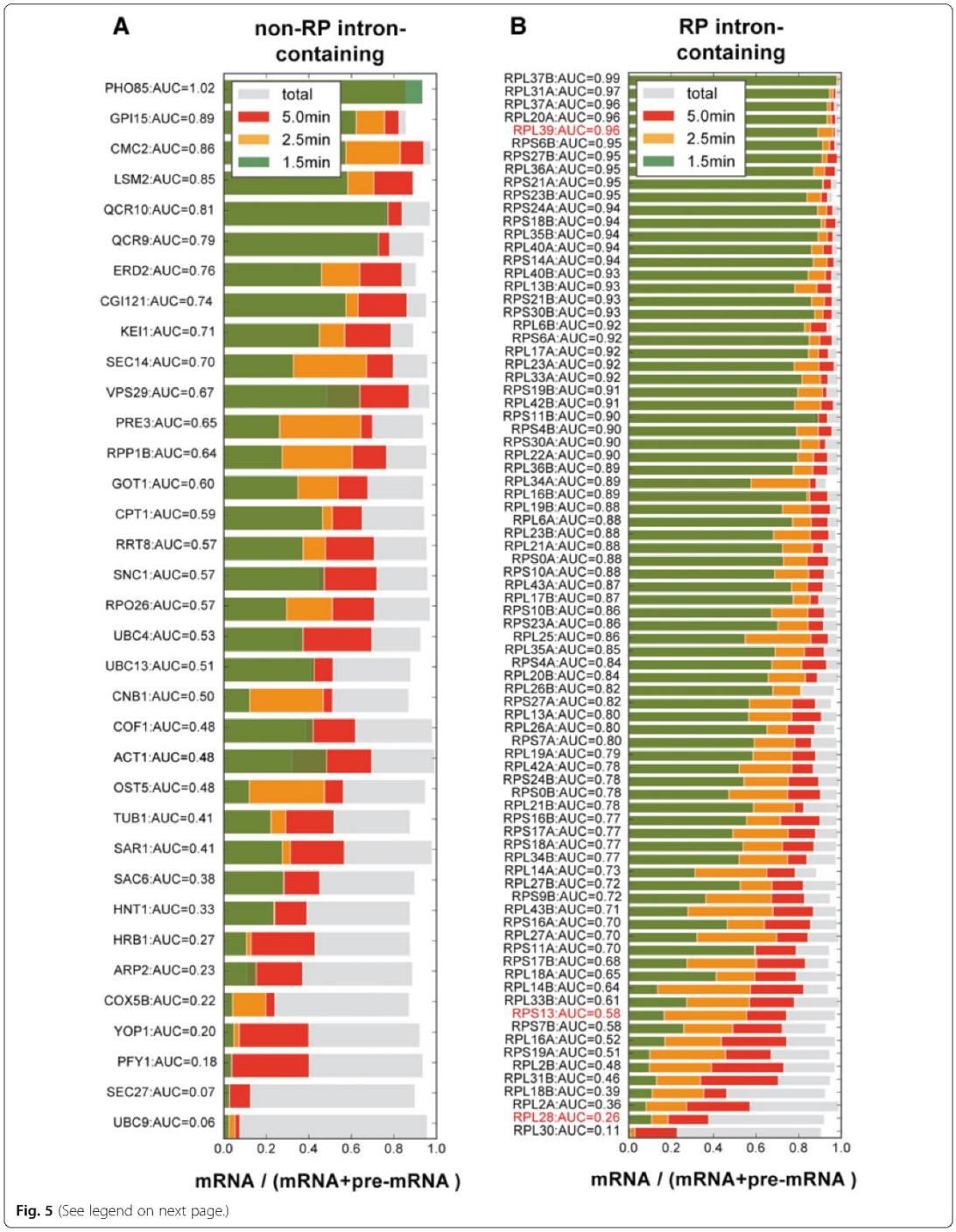


Fig. 5 (See legend on next page.)

(See figure on previous page.)
Fig. 5 The splicing speed and associated features. The mRNA proportions at 1.5 min, 2.5 min, 5.0 min and steady state for 35 non-ribosomal protein (*RP*) intron-containing genes (**a**) and 82 RP intron-containing genes (**b**). The proportion of mRNA is estimated using the probabilistic model described in Additional file 1: Data and Methods RNA-seq data, and the area under the curve (*AUC*) score denotes the splicing speed as defined in the "Methods" section. Faster splicing transcripts cluster at the top, and slower splicing transcripts at the bottom. *Red* transcripts were also validated by reverse-transcription quantitative polymerase chain reaction (Fig. 6). All four colored bars are overlapping

secondary structure of the intron. We looked for correlations between different transcript features and the relative speed of splicing. Analyzing the fastest-splicing and slowest-splicing thirds of transcripts, we noticed there was a marked difference in the behavior of intronic RP transcripts compared to that of non-RP intronic genes: for the intronic RP transcripts, a highly significant difference (Wilcoxon's test $p < 0.0003$) was found only with regard to the normalized secondary structure scores of RP introns, with the major contribution coming from the 5′ss to BP region (<1e04; see Fig. 7a). In the case of the non-RP transcripts, those that were spliced faster generally had less secondary structure at the 3′ss and a shorter exon 2 (Fig. 7b). All the feature comparisons are shown in Additional file 1: Figure S6 (non-RP transcripts) and Figure S7 (RP transcripts). The failure to see a significant effect of 5′ss, 3′ss and BP sequences in RP transcripts was likely due to the high similarity of these features.

The effect of intron secondary structure should be evident for paralogous RP genes that share highly related or identical exon sequences but have different intron sequences. Indeed, we found that some paralogs, such as RPL27A/B, RPS10A/B and RPL40A/B, that have introns with similar predicted secondary structure stabilities (ΔG values), had similar splicing speeds (Fig. 7c, left panel), whereas paralogs with introns that have different predicted ΔG values, such as RPL31A/B and RPS18A/B, had different splicing speeds (Fig. 7c, right panel).

Because the predicted secondary structure within RP introns was significantly correlated with the splicing speed, we further explored the intron sequences and found that several short base combinations were significantly correlated with splicing speed (Fig. 7d). Generally, fast splicing introns were enriched for adenosines, while slower splicing introns had a higher density of uridines. Surprisingly, the proportion of "A" and "U" in the introns was highly negatively correlated (Pearson's correlation coefficient < -0.75). Furthermore, we used the above features (all listed in Additional file 1: Tables S8 and S9) to predict the splicing speed by a random forest regression model. Figure 7e shows that the splicing speeds for RP pre-mRNAs could be well predicted by these features (Pearson's $R = 0.578$ between observed and predicted splicing speed).

Overall, the most significant features we have been able to identify that distinguish intron splicing rates for RP pre-mRNAs in budding yeast are the predicted secondary structure in the region between the 5′ss and the BP, and A or U density. In our data, slower splicing was associated with greater predicted secondary structure stability and U-richness in the intron. The effect of secondary structure was observed to be strongest for the set of highly expressed RP transcripts whose introns were mostly longer than average in budding yeast. The idea that secondary structure may have a role in splicing is not new [34–37], but previous work has mostly suggested that secondary structure in long introns in fact favors efficient splicing. Structure between the 5′ss and the BP has been proposed to be necessary for efficient splicing when this distance is greater than 200 nucleotides [37]. Work done on *RPS17B* by the Rosbash laboratory experimentally identified two complementary regions between the 5′ss and the BP that base paired to form a stem loop, thereby reducing the effective distance [38]. Mutations that disrupted this stem loop reduced the efficiency of splicing; compensatory mutations that restored the loop restored splicing efficiency. Further work by Rogic et al. [37] suggested that a specific structural arrangement was not required but that a very thermodynamically stable structure could slow splicing, possibly by masking splicing signals. Taken together with the results of Rogic et al. [37], our work indicates that efficient splicing of RP pre-mRNA transcripts requires an optimal amount of secondary structure between the 5′ss and the BP, with either too much or too little being detrimental. Furthermore, our results suggest that in the endogenous context, RP transcripts that splice slower are more often victims of too much structure rather than too little. A recent paper showed that fast splicing of reporter constructs correlated with low secondary structure around the splice sites [39]. For the RP genes we did not find a significant correlation between splicing speed and folding energies of splice sites (Additional file 1: Figure S7); However, this is presumably because the splice signals for these RP genes are very similar to the consensus sequences. We did, however, find a correlation between weak folding of the 3′ss sequence and efficient splicing for the non-RP genes analyzed here (Additional file 1: Figure S6), consistent with previous work [39].

Conclusions

4tU-seq is a powerful way of examining RNA processing kinetics and has many other potential applications. Our

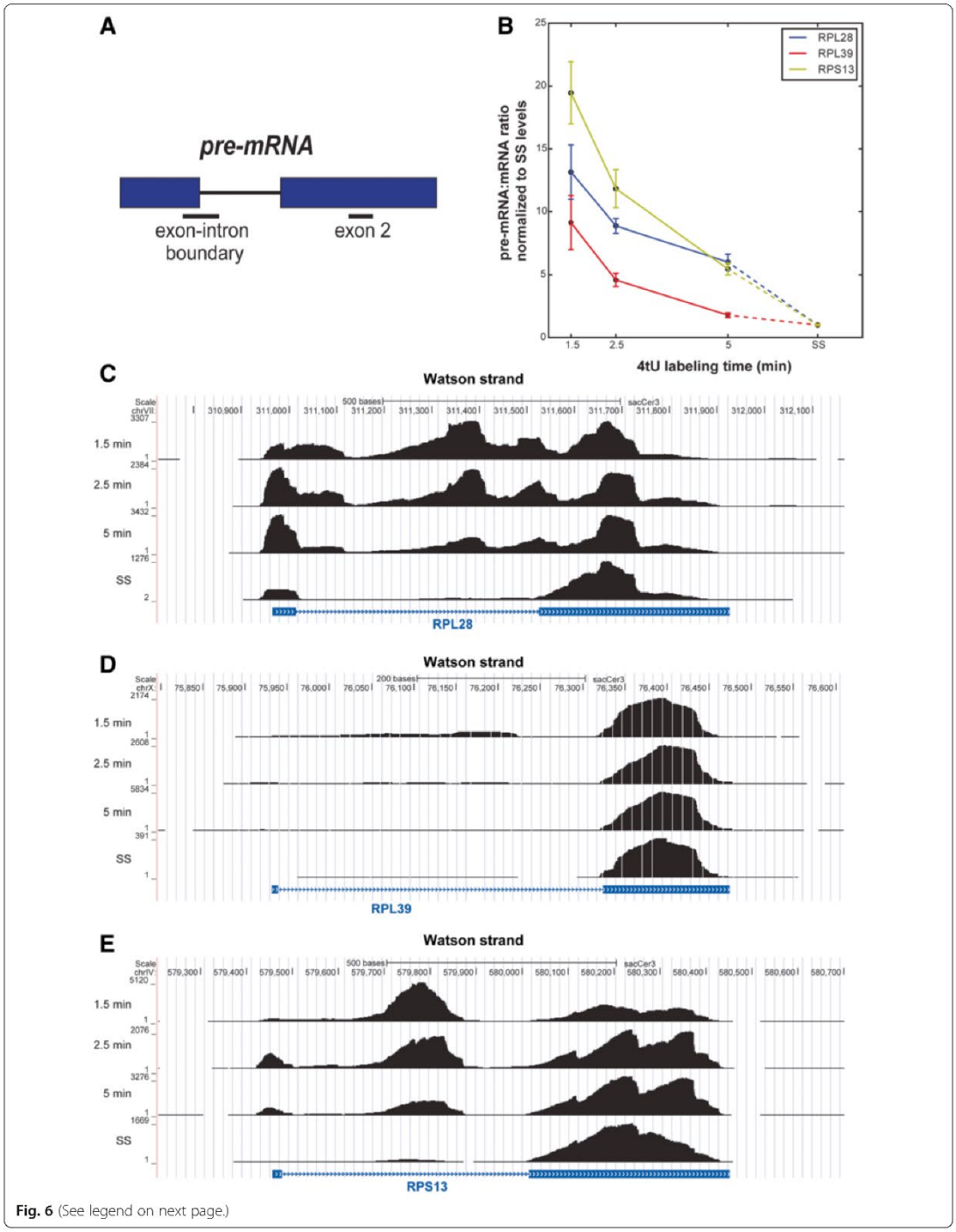


Fig. 6 (See legend on next page.)

(See figure on previous page.)
Fig. 6 Reverse transcription quantitative polymerase chain reaction (RT-qPCR) analysis of splicing status shows differences between transcripts. **a** Diagram showing the location of diagnostic amplicons. Exons are denoted by *blue boxes* and the intron is represented by a *black line*. Amplicons are indicated by lines below. Pre-mRNAs were detected using oligonucleotides that amplify the exon–intron boundary at the 5′ splice site. **b** Relative pre-mRNA levels of three transcripts, *RPL28, RPL39* and *RPS13*, analyzed by RT-qPCR, normalized relative to steady state (*SS*) levels. Data show how the level of each amplicon approaches the level detected at steady state as labeling time increases. Data were normalized to the levels of exon 2 and steady state to account for different RNA yields obtained at each labeling time. Different transcripts show different rates of splicing. **c**-**e** UCSC genome browser screen shots showing the change in distribution of reads at different labeling times (y-axis) for *RPL28, RPL39* and *RPS13*, with annotation below in *blue*. Exons are represented by *blue boxes* and intron indicated by a *blue line*. *SS* indicates steady-state levels, generated by sequencing total RNA

analysis at very short labeling times reveals unexpected complexity in RNA processing for several different families of RNAs. For lncRNAs, such as CUTs and SUTs, our data show that early abundances cannot be explained by simple first order kinetics, and quantitatively relates degradation of CUTs and SUTs to transcript length and predicted secondary structure.

Our analysis shows that to measure splicing kinetics in *S. cerevisiae* it is essential to recover the RNA after extremely short 4tU labeling times; many transcripts approach steady state by about 2.5 min. By measuring the speed of splicing of many different newly synthesized pre-mRNAs we were able to search for factors that could explain why some splice faster than others in vivo. For the 35 non-RP intronic genes for which we had adequate sequence coverage, we found a significant correlation between the secondary structure around the 3′ss and exon 2 length. This suggests that for these transcripts, selection of the 3′ss might be the rate-limiting step because it requires unwinding of RNA secondary structures. Differences were found even between intronic RP transcripts that have similar expression levels, the same gene annotation (Gene Ontology term), and similar lengths and strength of splice sites. Moreover, transcripts of certain highly related paralogous genes displayed differences in splicing speeds. Our results suggest that this difference in splicing kinetics is in part due to the secondary structure of the introns as well as the nucleotide composition. Within the RP gene subgroup, introns with a less favorable predicted secondary structure (less negative ΔG) spliced faster than those with more structure. One simple explanation for the observed trends is that it is more difficult for spliceosomes to assemble on introns with a stronger secondary structure and so more time is required to overcome this impediment. Another possibility is that structural re-arrangements required within the spliceosome to allow catalysis are costlier when there is more structure to contend with. Whether a transcript splices quickly or slowly carries over into its mRNA level at steady state, and it seems probable that regulating the secondary structure of the intron will in effect regulate the expression of the gene. Furthermore, in organisms with multi-intron genes and alternative splicing, an optimal degree of secondary structure could contribute to determining which introns get spliced in kinetic competition. It will be interesting to use 4tU-seq to study the kinetics of splicing under different metabolic and environmental conditions and to test the effects of different splicing factors on the speed of splicing.

Methods

Yeast strains and plasmids

The *ura3* point mutation in W303 (MAT**a**, ade2-1, ura3-1, his3-11, 15 trp 1-1, leu2-3, 112 can1-100) was corrected to create W303U. W303U was transformed with plasmid pAT1 (*FUI1* on the plasmid pRS425).

4tU-seq

Cultures were grown in Synthetic Defined −Ura −Leu medium to $OD_{600} = 0.8$, at which point 4tU was added to a final concentration of 500 μM. Cultures were maintained at 30 °C and were shaken throughout the experiment. Large (600 ml) cultures were used for 4tU-seq experiments. After the desired labeling time the cells were snap frozen by dropping the culture into a half volume of methanol in a large beaker sitting in dry ice to rapidly halt metabolism [10]. While the methanol slurry was still liquid, the frozen cells were pelleted by centrifugation at 3000 g for 3 min at 4 °C. RNA was isolated using a standard hot-phenol extraction. Biotinylation was performed on 2 mg of RNA as previously described [20], but only incubated for 15 min at 65 °C to reduce RNA degradation. The RNA was purified using 2 ml Zeba columns (ThermoFisher Scientific, Perth, UK), as per manufacturer's instructions, and ethanol precipitated. Biotinylated RNA was extracted using streptavidin C1 Dynabeads (ThermoFisher Scientific). We equilibrated 50 μl of beads in $Na_3PO_4TMgCl_2$ (0.2 M NaCl, 0.1 M sodium phosphate pH 6.8, 25 mM $MgCl_2$ and 0.4 % Tween) then blocked in $Na_3PO_4TMgCl_2$ and 1 μg/μl glycogen for 20 min. The biotinylated RNA was purified for 30 min with the Dynabeads at 4 °C in $Na_3PO_4TMgCl_2$. The beads were washed three times in $Na_3PO_4TMgCl_2$ and twice with TEN1000 (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH8 and 1 M NaCl). All bead block, incubation and wash volumes were 400 μl. The RNA was eluted twice with 50 μl of 0.7 M beta-mercaptoethanol for 5 min at room temperature before

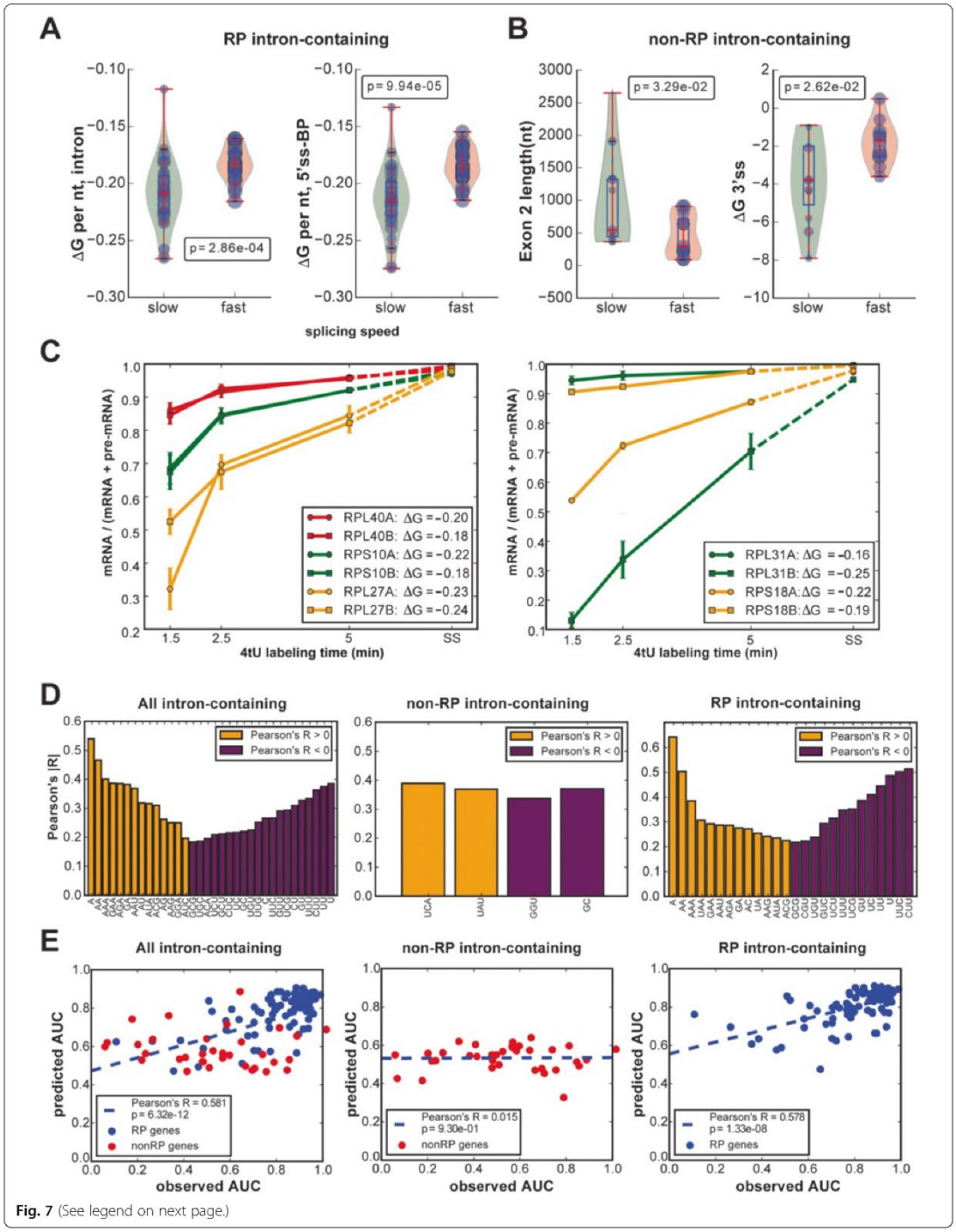


Fig. 7 (See legend on next page.)

(See figure on previous page.)
Fig. 7 Features associated with splicing speed and comparison of paralogs. **a** Comparison of secondary structure scores (ΔG; y-axis) for the fastest-splicing and slowest-splicing thirds of 82 ribosomal protein (*RP*) intron-containing genes (x-axis). The violin plots show the distribution of the features, and the *blue dots* represent individual RP genes, with dot size corresponding to the splicing speed. The *p*-value was obtained using Wilcoxon's test. **b** Comparison of exon 2 length and secondary structure at the 3′ splice site (3′ss) (y-axis) for 35 non-ribosomal protein (*non-RP*) intron-containing genes to splicing speed (x-axis). The violin plots show the distribution of the features, and the *blue dots* represent individual genes, with dot size corresponding to the splicing speed. The *p*-value was obtained using Wilcoxon's test. **c** The mRNA proportion changes of three pairs of paralogs, each pair of which show a similar splicing rate (*left panel*) and of three pairs of paralogs, each pair of which shows different splicing rates (*right panel*). The proportion of mRNA is estimated using the probabilistic model described in Additional file 1: Data and Methods from 4-thiouracil (*4tU*) data. The ΔG per nucleotide values (see "Methods" section) between the 5′ss and branch point are stated in the inset boxes. **d** Pearson's correlations between splicing speed and sequence patterns show the significantly correlated features ($p < 0.05$) to splicing speeds for all 117 intron-containing genes (*left panel*), 35 non-RP intron-containing genes (*middle panel*), and 82 RP intron-containing genes (*right panel*). The features are the occurrence of the specific base or bases in the intron. *Yellow* represents positive correlation with splicing speed and *purple* represents negative correlation. **e** Scatterplot of observed and predicted splicing speeds from the associated features. The features are listed in Additional file 1: Tables S8 and S9, and include secondary structures, splice site scores, intron length and exon length. The predictions are obtained by random forest regression with automatic feature selection. *AUC* area under the curve

being ethanol precipitated with 20 μg glycogen at −20 °C for at least 2 h and re-suspended in 10 μl DEPC-treated H_2O. To generate the 0 min or background RNA samples, we applied this protocol to cultures to which no 4tU had been added, which essentially provided an overview of RNAs that non-specifically bound to the streptavidin magnetic beads during the isolation procedure.

RNA-seq libraries were produced essentially as described previously [40]. Briefly, between 100 and 250 ng of (thiolated) RNA was fragmented in SuperScript Reverse Transcriptase buffer (ThermoFisher Scientific) for 5 min at 95 °C. Fragmented RNA was subsequently randomly primed as described [40]. Edinburgh Genomics (Edinburgh, UK) performed the 100-base pair paired-end sequencing using the Illumina HiSeq 2500 platform.

Differential analysis of transcript abundance

To determine what classes of transcripts were significantly enriched between two time points (1.5 and 5 min) and in the total RNA (Fig. 2a), we used DESeq2 [19]. Three biological replicates of each time point were used for these analyses. For each comparison, DESeq2 generated a list of transcript names (such as CUTs, SUTs and protein-coding genes) that were significantly over-represented in each time point. We used a *p*-value of 0.05 as threshold. We then counted the total number of transcripts from each class over-represented at each time point and divided that number by the total number of transcripts in each class. This showed us what fraction of each class was significantly enriched in the thiolated RNA.

Estimation of mRNA background levels in 4tU-seq data

To estimate the mRNA background levels in 4tU-seq data, we let the measured intronic FPKM percentage be α_0 and α_1 at 0 ("background") and 1.5 min, respectively. We let the intronic FPKM percentage be β for signal; this value was unknown but we know from theoretical considerations that $\beta \leq 0.5$. We denoted the background and the signal fractions of RNA mapped to intron-containing genes at 1.5 min as m and n respectively. This gave us $\alpha_0 \times m + \beta \times n)/(m + n) = \alpha_1$, so:

$$m/n = (\beta - \alpha_1)/(\alpha_1 - \alpha_0) \leq (0.5 - \alpha_1)/(\alpha_1 - \alpha_0)$$

Therefore, the proportion of background reads over the total was bounded by $[(m/n)/(1 + m/n)]$. Substituting the observed values for α_0 and α_1, we obtained the 9.1 % upper bound as reported in the main text.

RT-qPCR

RT-qPCR was carried out as described previously [10] using oligonucleotides listed in Additional file 1: Table S5.

Processing of raw sequencing data

Sequencing was performed on a HiSeq 2500 by Edinburgh Genomics. Raw fastq files were demultiplexed using pyBarcodeFilter version 2.3.3 from the pyCRAC tool suite, version 1.2.2.4 [41]. Quality trimming and removal of 3′ adapter sequences (5′- AGATCGGAAGAGCACACG-3′) was performed using Flexbar, version 2.4 [42]. RNA-seq reads were aligned to the yeast genome by STAR [43]. Counts for annotated genomic features were generated using pyCRAC [41], in-house python scripts and genomic feature files (GTF) from ENSEMBL, version R64-1-1.75. Coordinates for CUTs and SUTs were obtained from Xu et al. [26]. The python scripts used for this study are available upon request.

Quantification of splicing ratio and splicing speed

The splicing ratio, that is, the mRNA proportion, was then calculated from the aligned reads either by direct methods or using a probabilistic method (see Additional file 1: Data and Methods); the code for the probabilistic method is available at https://github.com/huangyh09/diceseq. To combine the measurements of pre-mRNA and mRNA abundance at 1.5, 2.5 and 5 min into a single

measure of splicing speed, we considered the AUC as follows:

$$AUC = \frac{(R_{2.5} + R_{1.5})/2 + 2.5 \times (R_{5.0} + R_{2.5})/2}{3.5 \times R_{SS}} \quad (1)$$

where R_i is the splicing ratio, that is, mRNA/(mRNA + pre-mRNA), at time point i (in minutes), estimated as the average of the posterior mean for the three replicates. R_{SS} represents the merged splicing ratio at the steady state. Therefore, the larger the AUC, the more efficient the splicing process.

Intron-containing RP gene annotation and features

The lengths and sequences of various features were obtained from SGD (http://www.yeastgenome.org/). BP locations were obtained from the Yeast Intron Database [44]. The scores assigned to the 5′ss, 3′ss and BS were obtained using the method described by Crooks et al. [45]. The free-energy of the predicted secondary structure was calculated using quickfold (http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold). This value was then divided by the number of bases in the intron (or part of the intron) to give the secondary structure per base (a ΔG value and therefore negative). The more negative the ΔG value, the more structured the intron is predicted to be.

Data access

Raw (fastq) and processed sequencing data can be downloaded from the NCBI Gene Expression Omnibus repository [GEO: GSE70378]. All raw read count and feature data can be found in Additional file 1: Tables S6 to S14.

Ethical approval

No ethical approval was required for this study.

Additional file

Additional file 1: Tables S1–S14; figures S1–S7; and Supplementary Methods and References. (ZIP 6028 kb)

Abbreviations

3′ss: three prime splice site in intron; 4tU-seq: 4-thiouracil labeling and sequencing; 5′ss: five prime splice site in intron; BP: branch point sequence; CI: confidence interval; CUT: cryptic unstable transcript; ETS: external transcribed spacer ; FPKM: fragments per kilobase per million reads ; ITS: internal transcribed spacer ; lncRNA: long non-coding RNA; mRNA: messenger RNA; miRNA: microRNA; ncRNA: non-coding RNA; non-RP: non-ribosomal protein; RNA-seq: RNA sequencing; RP: ribosomal protein; siRNA: small interfering RNA; snRNA: small nuclear RNA; snoRNA: small nucelolar RNA; SS: steady state; SUT: stable unannotated transcripts; tRNA: transfer RNA; XUT: Xrn1-dependent unstable transcript.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed to planning the experiments and computational procedures. JR, DB, RH and SG carried out the experiments. YH, GS and SG performed the bioinformatics and computational analyses of the sequencing data. All authors contributed to writing the manuscript and approved the final manuscript.

Acknowledgements

This work was supported by grants from the Wellcome Trust to JR (093853), JB (087551), and SG (091549), and the Wellcome Trust Centre for Cell Biology core grant (092076). GS acknowledges support from the European Research Council under grant MLCS306999. YH is supported by the University of Edinburgh Principal Career Development scholarship. JB is the Royal Society Darwin Trust Research Professor. Next-generation sequencing was carried out by Edinburgh Genomics, The University of Edinburgh, Edinburgh, UK. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1).

Author details

[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK. [2]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK. [3]Centre for Synthetic and Systems Biology (SynthSys), University of Edinburgh, Edinburgh EH9 3BF, UK. [4]Present Address: Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK.

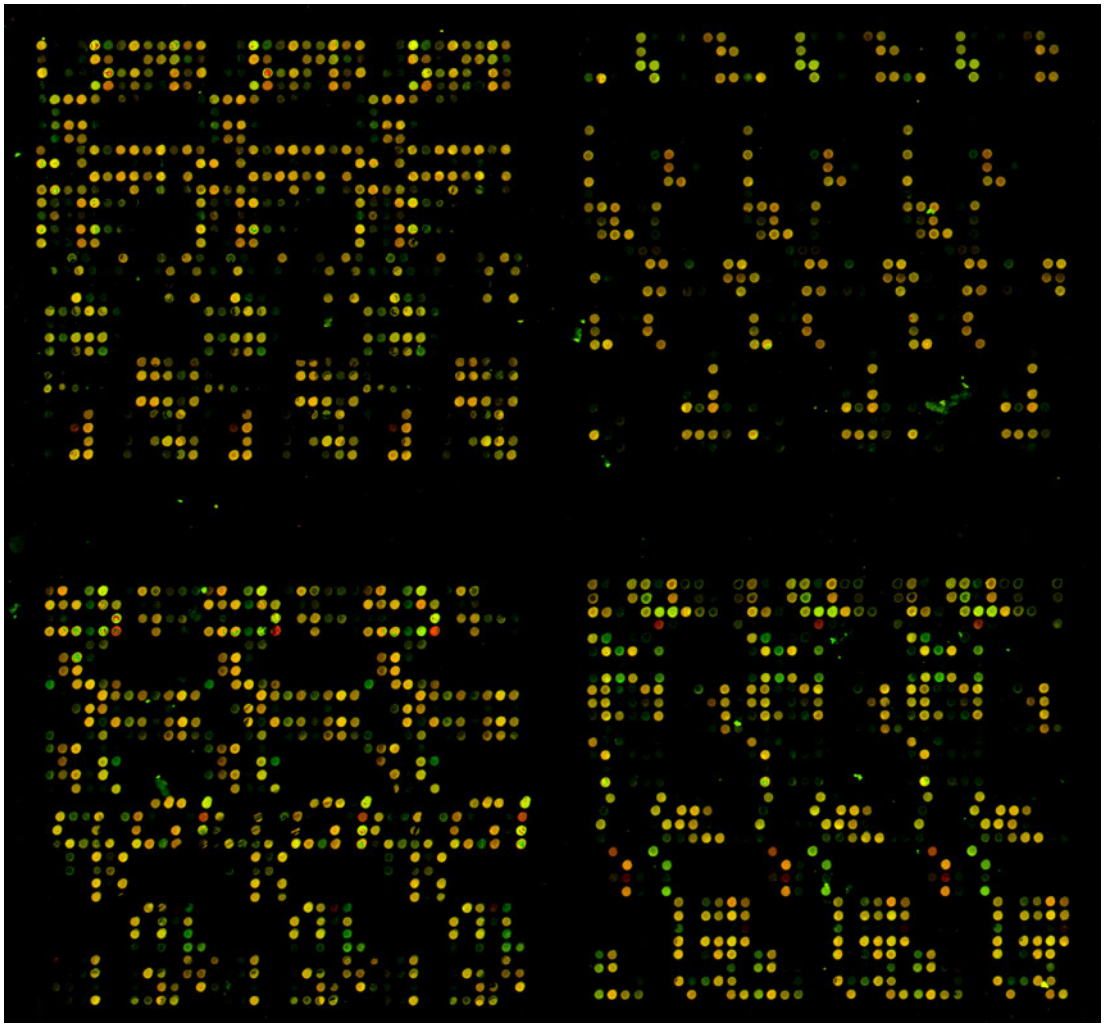
Received: 4 October 2015 Accepted: 30 November 2015
Published online: 17 December 2015

References

1. Jensen TH, Jacquier A, Libri D. Dealing with pervasive transcription. Mol Cell. 2013;52:473–84.
2. Tuck AC, Tollervey D. RNA in pieces. Trends Genet. 2011;27:422–32.
3. Yamashita A, Shichino Y, Yamamoto M. The long non-coding RNA world in yeasts. Biochim Biophys Acta. 2015. doi:10.1016/j.bbagrm.2015.08.003.
4. Tudek A, Candelli T, Libri D. Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? Biochimie. 2015;117:28–36.
5. van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature. 2011;475:114–7.
6. Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3. doi:10.1101/cshperspect.a003707.
7. Herzel L, Neugebauer KM. Quantification of co-transcriptional splicing from RNA-seq data. Methods. 2015;85:36–43.
8. Singh J, Padgett RA. Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol. 2009;16:1128–33.
9. Huranová M, Ivani I, Benda A, Poser I, Brody Y, Hof M, et al. The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. J Cell Biol. 2010;191:75–86.
10. Alexander RD, Barrass D, Dichtl B, Kos M, Obtulowicz T, Robert M-C, et al. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. RNA. 2010;16:2570–80.
11. Gray JM, Harmin DA, Boswell SA, Cloonan N, Mullen TE, Ling JJ, et al. SnapShot-Seq: a method for extracting genome-wide, in vivo mRNA dynamics from a single total RNA sample. PLoS One. 2014;9:e89673.
12. Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA. 2008;14:1959–72.
13. Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol. 2011;7:458.
14. Neymotin B, Athanasiadou R, Gresham D. Determination of in vivo RNA kinetics using RATE-seq. RNA. 2014;20:1645–52.
15. Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, et al. Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell. 2013;155:1075–87.
16. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol. 2005;23:232–7.

17. Wagner R, de Montigny J, de Wergifosse P, Souciet JL, Potier S. The ORF YBL042 of *Saccharomyces cerevisiae* encodes a uridine permease. FEMS Microbiol Lett. 1998;159:69–75.
18. Blondel M-O, Blondel MO, Morvan J, Dupré S, Urban-Grimal D, Haguenauer-Tsapis R, et al. Direct sorting of the yeast uracil permease to the endosomal system is controlled by uracil binding and Rsp5p-dependent ubiquitylation. Mol Biol Cell. 2003;15:883–95.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
20. Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, et al. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res. 2012;22:2031–42.
21. Burger K, Mühl B, Kellner M, Rohrmoser M, Gruber-Eber A, Windhager L, et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. 2013;10:1623–30.
22. Rasmussen TP, Culbertson MR. The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in *Saccharomyces cerevisiae*. Mol Cell Biol. 1998;18:6885–96.
23. Steinmetz EJ, Conrad NK, Brow DA, Corden JL. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature. 2001;413:327–31.
24. Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. Distinct pathways for snoRNA and mRNA termination. Mol Cell. 2006;24:723–34.
25. Qu LH, Henras A, Lu YJ, Zhou H, Zhou WX, Zhu YQ, et al. Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. Mol Cell Biol. 1999;19:1144–58.
26. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009;457:1033–7.
27. Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, et al. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. bioRxiv. 2015; 025585. http://biorxiv.org/content/biorxiv/early/2015/08/26/025585.full.pdf.
28. Colin J, Libri D, Porrua O. Cryptic transcription and early termination in the control of gene expression. Genet Res Int. 2011;2011:653194.
29. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. 2012;22:1616–25.
30. Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, et al. Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell. 2012;48:409–21.
31. Bousquet-Antonelli C, Presutti C, Tollervey D. Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell. 2000;102:765–75.
32. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7:1009–15.
33. Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YYC, Codlin S, et al. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. Genome Res. 2014;24:1169–79.
34. Awan AR, Manfredo A, Pleiss JA. Lariat sequencing in a unicellular yeast identifies regulated alternative splicing of exons that are evolutionarily conserved with humans. Proc Natl Acad Sci U S A. 2013;110:12762–7.
35. Meyer M, Plass M, Pérez-Valle J, Eyras E, Vilardell J. Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. Mol Cell. 2011;43:1033–9.
36. Parker R, Patterson B. 9 - Architecture of fungal introns: implications for spliceosome assembly. In: Dudock MIS, editor. Molecular biology of RNA. San Diego, CA: Academic Press, Inc.; 1987. p. 133–149. [Molecular Biology of RNA].
37. Rogic S, Montpetit B, Hoos HH, Mackworth AK, Ouellette BF, Hieter P. Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in *Saccharomyces cerevisiae*. BMC Genomics. 2008;9:355.
38. Goguel V, Rosbash M. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell. 1993;72:893–901.
39. Zafrir Z, Tuller T. Nucleotide sequence composition adjacent to intronic splice sites improves splicing efficiency via its effect on pre-mRNA local folding in fungi. RNA. 2015;21:1701–18.
40. Hector RD, Burlacu L, Aitken S, Le Bihan T, Tuijtel M, Zaplatina A, et al. Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. Nucleic Acids Res. 2014;42:12138–54.
41. Webb S, Hector RD, Kudla G, Granneman S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol. 2014;15:R8.
42. Dodt M, Roehr JT, Ahmed R, Dieterich C. FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. Biology (Basel). 2012;1:895–905.
43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
44. Spingola M, Grate L, Haussler D, Ares M. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA. 1999;5:221–34.
45. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

5 eIF4G: Translation Contacts Splicing



Scan of a splicing microarray;
See thesis introduction for design details. For this microarray the mutant strain's RNA has been reverse transcribed and labelled with Cy3 (red) and the wild type with Cy5 (green). The mutant splices poorer; introns are more abundant, conversely mature mRNAs are less so. With this labelling strategy, the redder spots are probes corresponding to introns, the greener to exon/exon junction probes. Probes that do not have different abundances in the two samples are a combination of green and red and so appear orange, often these are probes to the exons. In this way the splicing status of all intron containing transcripts in one sample versus another can be determined.

5.1 Research Article

This chapter is based on the peer-reviewed publication: Kafasla, P., Barrass, J.D., Thompson, E., Fromont-Racine, M., Jacquier, A., Beggs, J.D., Lewis, J., 2009. Interaction of yeast eIF4G with spliceosome components. RNA Biol 6, 563–574.

Figures from this publication will be referred to as P.Figure followed by the number.

5.2 Aim

Metazoan eukaryotic translation initiation factor eIF4G is involved in RNA metabolism as well as translation. This study investigated if this is also true for eIF4G of *S. cerevisiae*. Yeast has two copies of the gene encoding eIF4G, the paralogues Tif4631p and Tif4632p (Goyer et al., 1993).

5.2.1 eIF4G/Tif4631p/Tif4632p

eIF4G's first recorded function was as part of the translation initiation factor eIF4F (Grifo et al., 1983). Since its initial identification, several roles in RNA metabolism have been proposed (see Figure 1 for known interactions, with emphasis on those discussed in the publication).

5.3 Experiments

5.3.1 Subcellular localisation

TAP-tagged Tif4631p was found to be evenly distributed between the nucleus and cytoplasm, whereas Tif4632p was mainly in the cytoplasm with only a little in the nucleus (P.Figure 1). This observation suggests that the eIF4G may indeed have more than a translation role, and that the two versions could have a core translation function but different additional functions.

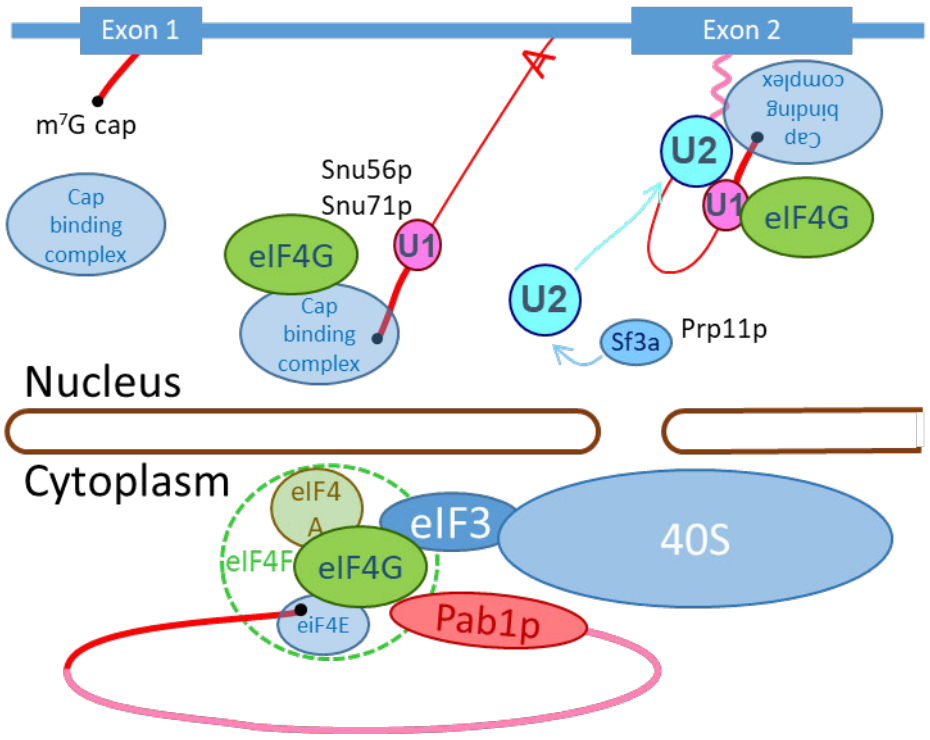


Figure 1: Roles of eIF4G
eIF4G (Tif4631p and Tif4632p in yeast), has roles in both the cytoplasm and the nucleus. Indicated are some of the proteins and complexes it interacts with (Das and Das, 2016). eIF4G, eIF4A and eIF4E are all part of the eIF4F complex.

A ClustalW (Thompson et al., 1994) and nuclear localisation probability (NucPred) (Brameier et al., 2007) plot is shown on the next page. Both Tif4631p and Tif4632p are predicted to be nuclear with reasonable confidence. They have a NucPred score of 0.98 and 0.92 respectively and, although Tif4631p is more likely nuclear than its paralogue, they both have a >88 % likelihood of being present in the nucleus, based on correlation between *in vivo* observations and the score.

From the alignment on the next page the homology is highest from amino acid 508 in Tip4631p and its counterpart in Tif4632p, amino acid 473, and conserved from about 430 (about 400 in Tif4632p). This region includes most of the binding sites identified, see later.

```
Tif4631p 1    MTDETAHPTQSASKQESAALKQTGDDQQESQQQRGYTNYNNGSNYTQKKPYNSNRPHQQR
Tif4632p 1    MTDQRGPPPPHPQQANGYKKFPPHDNQYSGANNSQPNNHYNENLYSAREPHN-NKQYQSK
consensus     MTD    P               D Q          N  N   Y    P N N   Q

Tif4631p 61   GGKFGPNRYNNRGNYNGGGSFR------GGHMGANSSN------VPWTGYYNNYPVYYQP
Tif4632p 60   NGKYGTNKYNNRNNSQGNAQYYNNRFNNGYRLNNNDYNPAMLPGMQWPANYYAPQMYYIP
consensus      GK G N YNNR N  G           G     N  N        W   Y     YY P

Tif4631p 109  QQMAAAGS-----APANPIPVEEKSPVPTKIEITTKSGEHLDLKEQHKAKLQSQ-ERSTV
Tif4632p 120  QQMVPVASPPYTHQPLNTNPEPPSTPKTTKIEITTKTGERLNLKKFHEEKKASKGEEKND
consensus     QQM    S      P N  P     P  TKIEITTK GE L LK  H  K  S  E

Tif4631p 163  SPQPESKLKETSDSTSTSTPTPTPSTND--SKASSEENISEAEKTRRNFIEQVKLRKAAL
Tif4632p 180  GVEQKSKSGTPFEKEATPVLPANEAVKDTLTETSNEKSTSEAENTKRLFLEQVRLRKAAM
consensus          SK         T          D     S E   SEAE T R F EQV LRKAA

Tif4631p 221  EKKRKEQLEGSSGNNNIPMKTTPENVEEKGSDKPEVTEKTKPAEEKSAEPEVKQETPAEE
Tif4632p 240  ERKKNGLIS----------ETE----KKQETSNHDNTDTTKPNS--VIESEPIKEAPKPT
consensus     E K                T               T  TKP      E E   E P

Tif4631p 281  GEQGEKGQIKEESTPKVLTFAERLKLKKQQKEREEKTEGKENKEVPVQEETKSAIESAPV
Tif4632p 284  G----------EANEVVIDGKSGASVKTPQH-----VTGSVTKSVTFNEP-EN--ESSSQ
consensus     G          E    V         K  Q        G   K V   E      ES

Tif4631p 341  PPSEQVKEETEVAETEQSNIDESATTPAIPTKSDEAEAEVEAEAGDAGTKIGLEAEIETT
Tif4632p 326  DVDELVKDD------DTTEISDTTGGKTVNKSDDETINSVITTEEN--T-----VK-ETE
consensus        E VK           I             DE    V        T        ET

Tif4631p 401  TDETDDGTNTVSHILNVLKDATPIEDVFSFNYPEGIEGPDIKYKKEHVKYTYGPTFLLQF
Tif4632p 372  PSTSDIEMPTVSQLLETLGKAQPISDIYEFAYPENVERPDIKYKKPSVKYTYGPTFLLQF
consensus        D    TVS  L  L  A PI D   F YPE  E PDIKYKK  VKYTYGPTFLLQF

Tif4631p 461  KDKLNVKADAEWVQSTASKIVIPPGMGRGNRSRDSGRFGNNSSRGHDFRNTSVRNMDDRA
Tif4632p 432  KDKLKFRPDPAWVEAVSSKIVIPPHIAR-NKPKDSGRFGG------DFRSPSMRGMDHTS
consensus     KDKL    D  WV    SKIVIPP   R N   DSGRFG       DFR  S R MD

Tif4631p 521  NSRTSSKRRSKRMNDDRRSNRSYTSRRDRERGSYRNEEKREDDKPKEEVAPLVPSANRWV
Tif4632p 485  SSRVSSKRRSKRMGDDRRSNRGYTSRKDREK----AAEKAEEQAPKEEIAPLVPSANRWI
consensus      SR SSKRRSKRM DDRRSNR YTSR DRE       EK E   PKEE APLVPSANRW

Tif4631p 581  PKFKSKKTEKKLAPDGKTELLDKDEVERKMKSLLNKLTLEMFDAISSEILAIANISVWET
Tif4632p 541  PKSRVKKTEKKLAPDGKTELFDKEEVERKMKSLLNKLTLEMFDSISSEILDIANQSKWED
consensus     PK   KKTEKKLAPDGKTEL DK EVERKMKSLLNKLTLEMFD ISSEIL IAN S WE

Tif4631p 641  NGETLKAVIEQIFLKACDEPHWSSMYAQLCGKVVKELNPDITDETNEGKTGPKLVLHYLV
Tif4632p 601  DGETLKIVIEQIFHKACDEPHWSSMYAQLCGKVVKDLDPNIKDKENEGKNGPKLVLHYLV
consensus      GETLK VIEQIF KACDEPHWSSMYAQLCGKVVK L P I D  NEGK GPKLVLHYLV

Tif4631p 701  ARCHAEFDKGWTDKLPTNEDGTPLEPEMMSEEYYAAASAKRRGLGLVRFIGFLYRLNLLT
Tif4632p 661  ARCHEEFEKGWADKLPAGEDGNPLEPEMMSDEYYIAAAAKRRGLGLVRFIGYLYCLNLLT
consensus     ARCH EF KGW DKLP  EDG PLEPEMMS EYY AA AKRRGLGLVRFIG LY LNLLT

Tif4631p 761  GKMMFECFRRLMKDLTDSPSEETLESVVELLNTVGEQFETDSFRTGQATLEGSQLLDSLF
Tif4632p 721  GKMMFECFRRLMKDLNNDPSEETLESVIELLNTVGEQFEHDKFVTPQATLEGSVLLDNLF
consensus     GKMMFECFRRLMKDL   PSEETLESV ELLNTVGEQFE D F T QATLEGS LLD LF

Tif4631p 821  GILDNIIQTAKISSRIKFKLIDIKELRHDKNWNSDKKDNGPKTIQQIHEEEERQRQLKNN
Tif4632p 781  MLLQHIIDGGTISNRIKFKLIDVKELREIKHWNSAKKDAGPKTIQQIHQEEEQLRQKKNS
consensus       L  II    IS RIKFKLID KELR  K WNS KKD GPKTIQQIH EEE  RQ KN

Tif4631p 881  SRSNSRRTNNS---SNRHS-FRRDAPPASKDSFITTRTYSQRNSQRAPPPKEEPAAPTST
Tif4632p 841  QRSNSRFNNHNQSNSNRYSSNRRNMQNTQRDSFASTKTGSFRNNQRNARKVEEVSQAP--
consensus      RSNSR  N     SNR S  RR       DSF  T T S RN QR     EE

Tif4631p 937  ATNMFSALMGESDDEE
Tif4632p 899  RANMFDALMNNDGDSD
consensus       NMF ALM    D

(non-nuclear) negative |||||||||||||||||||||||||| positive (nuclear)
```

5.3.2 Interactions with snRNPs

TAP-tagged Tif4631p and Tif4632p were immunoprecipitated from whole cell extracts. Spliceosomal snRNAs associated with either of the Tif proteins were identified by northern blots (P.Figure 2). Increasing the salt concentration when washing the immunoprecipitation gives an indication of the strength of interaction. U1 is reliably pulled down even under high salt by both eIF4G paralogues. U6 also seems to bind strongly, particularly with Tif4632p, possibly even more strongly than U1. Tif4632p's associations are also generally more resistant to salt. U2 is poorly detected so it is not clear what form of the spliceosome is being detected, a possibility is the pre-B form as all U snRNPs are detected. As U1 is very strongly detected it is likely that spliceosomes prior to the pre-B could also be associated with eIF4G (see Figure 1). The low abundance of U2 in the pulldowns could be an indication that the interaction with eIF4G is indirect.

5.3.3 Interactions with Proteins

Prp11p and Snu71p were previously identified by a yeast 2-hybrid screen as interactors with the Tif proteins, although Snu71p only with Tif4631p. GST fusions of these two proteins were made and used to capture *in vitro* transcribed then *in vitro* translated ^{35}S labelled Tif4631/2p, P.Figure 3. Prp11p did indeed pull down the two Tif proteins, and Snu71 also pulled down both proteins, albeit at a lower level. Pull down of Tif4631p was examined further and proved to be somewhat salt resistant for both Prp11p and Snu71p. The interaction with Prp11p was only slightly sensitive to RNases.

This analysis was extended to find the minimal regions of Tif4631p that interact with Prp11p, Snu71p and the cap binding complex via Cbc2p (its alias, Mud13p, was used in the publication), P.Figure 4. The method of analysis was essentially the same as described above, using ^{35}S labelled, but truncated Tif proteins and where some amino acids were replaced by alanines. This, along with interacting domains already known, is summarised in Figure 2, Table 1 and P.Figure 4F. There is some confusion in the legend in the paper, where it is implied that full length Tif4631p and truncated Tif4632p was used. The

truncations were done in Tif4631p and some of them were repeated in Tif4632p even though they are not shown.

Three RNA binding regions (described as RRMs in Das and Das, (2016)), have been located in Tif4641p (Berset et al., 2003) (Park et al., 2011), one at the N terminus (1-81), near the C terminus (883-952), and in the middle (492-539). They were not identified in Tif4632p, but they were mentioned as probable, without presenting the evidence (Berset et al., 2003). The conservation in those regions indicates at least the latter two are likely to exist in the paralogue. The N terminal RRM probably interacts with poly(A) tails (Park et al., 2011)). The middle one almost precisely overlaps the region identified as associating with Prp11p. The interaction with Prp11p is somewhat confusing when compared with the U2 snRNP pulldown. Prp11p is described in the publication as a U2 protein, so U2 would be expected to be more efficiently recovered than it is. The binding with Prp11p could be through RNA, but the interaction between Prp11p and eIF4G was not RNase sensitive, making this unlikely. A possible, although slightly unsatisfactory explanation, is that Prp11p is actually part of SF3a, and while it joins the pre-mRNA with U2 it is soon dissociated by Prp2p (Ohrt et al., 2012), but remains in association with the spliceosome, still found in mass-spectrometry studies until the P complex (Will and Lührmann, 2011). So, it is possible that the pull down reflects SF3a's later associations.

The middle to C-terminus portion of eIF4G is rich in binding sites (Figure 2), only the Pab1p and the terminal RNA binding domains are not in this region. These sites overlap which might indicate mutually exclusive interactions. The Snu71p binding site overlaps with both the CBC and MIF4G domain (a protein-protein interaction domain first identified in eIF4G), this site is the binding site for both eIF4A and eIF3. While Snu71/eIF interactions will happen at different stages in the life of an RNA, the CBC and U1 will both be present near the site of transcription. If they are mutually exclusive, there will be a handover of eIF4G from CBC to U1, as hinted at in Figure 1. The Prp11p and the middle RNA binding site also occupy the same region. Only one further study

examined this region and proved its importance to viability but did not examine its RNA binding (Park et al., 2011).

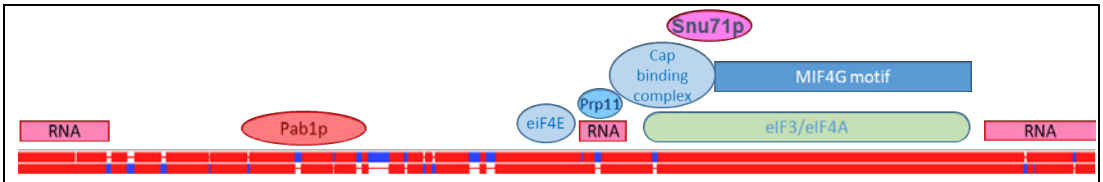


Figure 2: Binding Regions of *S. cerevisiae* eIF4G
The binding sites (and motif locations), are mapped onto an alignment of Tif4631p and Tif4632p. Most sites are clustered in a conserved region from the middle to C-terminus. The ClustalW alignment is at the bottom; Tif4631p - upper bar and Tif4632p - lower bar, matches are thick red bars, insertions blue and gaps thin red sections.

Table 1: Tif4631p Binding site locations

Binding site / Motif location	Location in Tip4631p
RNA (binds RNA near the poly-A-tail)	1 to 81
Pab1p	188 to 299
eIF4E	441 to 490
RNA	492 to 539
Prp11p	494 to 529
Cap Binding Complex (CBC)	522 to 612
eIF3*/eIF4A	552 to 847
Snu71p	567 to 647
MIF4G motif	616 to 850
RNA	883 to 952 (end)

*Interaction with eIF3 may not be present in *S. cerevisiae*

5.3.4 eIF4's Protein Interactions Affect Splicing

An *in vitro* splicing assay was performed in the presence of increasing amounts of the regions of the Tif proteins that interact with Prp11p (453-647 in Tif4631p), and both Prp11p & Snu71p (424-609 in Tif4632p), P.Figure 5. Adding these peptides reduced splicing. This could be due to the peptides sequestering Prp11p, Snu71p and/or binding the RNA directly.

5.3.5 Splicing is Affected by Deletion

Microarrays, as discussed in the thesis introduction, were performed on tif4631-Δ and tif4632-Δ strains P.Figure 6A for design, B and C for validation). No splicing defect was discovered in tif4632-Δ. Only 6 transcripts were identified as being affected by the loss of tif4631p (P.Table 2). The low number of affected transcripts is not surprising as the two paralogues are redundant. Tif4631p seems to be the more abundant protein (Clarkson et al., 2010), so the defect is probably related to a lower abundance protein failing to fully substitute for a more abundant one. There is nothing obviously differentiating the affected transcripts from any other intron-containing transcripts that would indicate that eIF4G is required specifically for their splicing. All six are RPGs and each gene is one of a pair of paralogues (the other copy unaffected), so intron retention could be a factor in their regulation. Many RPGs exist in duplicate and many spliced transcripts are RPGs, so there is not much that can be concluded from the small set identified. However, due to the RPGs being affected the result is consistent with a splicing defect. Splicing microarrays used to analyse exosome deletion strains produce a more eclectic collection of affected transcripts with both positive and negative effects on the efficiency of splicing (Figure 3).

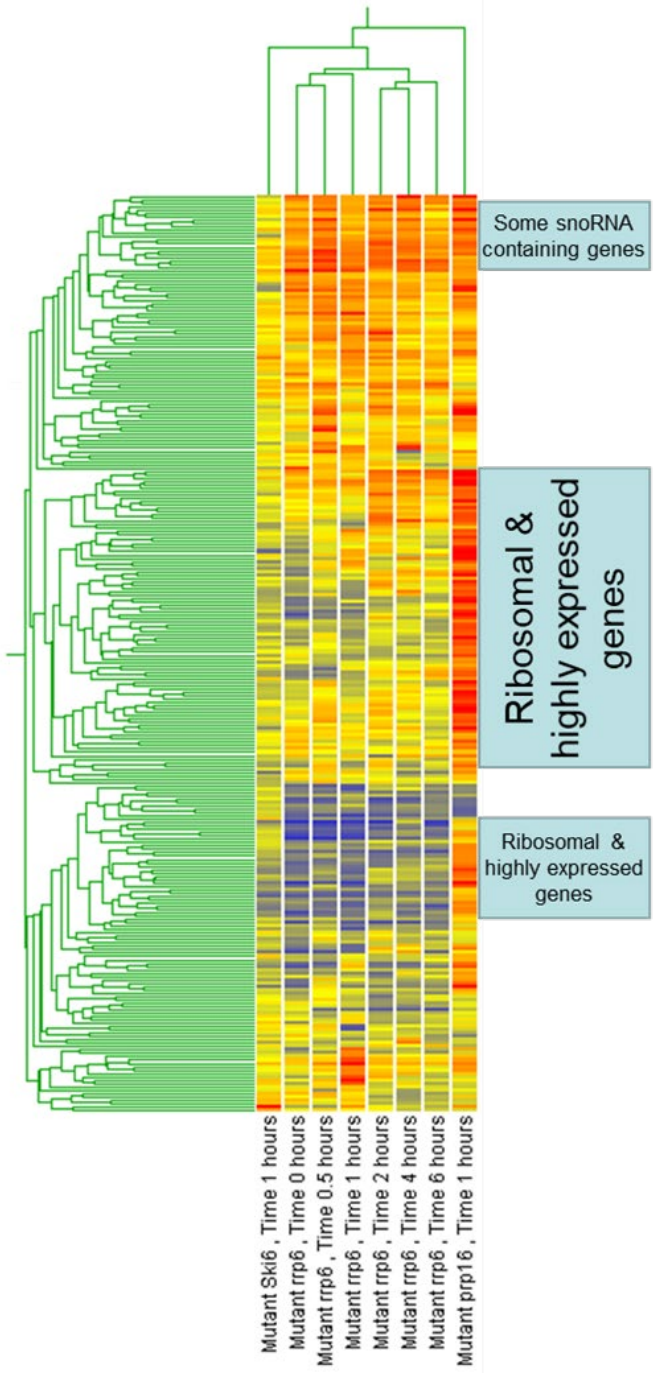


Figure 3: Degradation affects Splicing
Hierarchical clustering of microarray data comparing degradation factor deletion mutant strains Δski6 and Δrrp6 with the temperature sensitive splicing mutant strain prp16-2 (see chapter 7 for more work on this factor). Red indicates poorer splicing of that gene (in rows), blue improved splicing. Columns are different strains and times after shift to non-permissive temperature (labelled at the bottom).
While both Ski6p and Rrp6p are members of the exosome, Rrp6p is nuclear and its deletion confers temperature sensitivity. The exosome deletion strains either have no effect on splicing or improve it, possibly as more time is allowed for splicing before degradation. A few are poorer spliced, mainly transcripts containing a snoRNA, presumably the exosome is involved in their release from the transcript. None of the transcripts identified as poorer spliced in Δtif4631 are poorer spliced in Δski6 or Δrrp6 but they are poorer spliced in prp16-2.

5.4 Re-interpretation

Since this paper was published there have been reports of eIF4G's involvement in RNA surveillance (Das et al., 2014) as well as splicing and translation. In review (Das and Das, 2016) eIF4G has been described as a

general adapter protein for RNA metabolism. The publication in this chapter is still the primary reference for TIF proteins and splicing and has contributed to this view.

Given the proposed connection to surveillance (Das et al., 2014), where several RNAs were stabilised by deletion of *TIF4631*, it is possible that the results showing that the TIF4631Δ strain has poorer splicing is, in fact, due to stabilisation of unspliced precursors that would otherwise have undergone decay. This would imply that the interactions between eIF4G and splicing factors are to check correct association of these factors with the RNA. With the paucity of articles published on eIF4G and splicing since this publication, it is currently impossible to tell if this is the mechanism by which eIF4G affects splicing. From the microarray and *in vitro* splicing assay, a splicing defect is more likely than reduced degradation, but the available evidence is far from conclusive.

There are differences between the two paralogues, as found in this publication. Tif4632p is more cytoplasmic, prefers U6 snRNP to U1 and more stably associates with the snRNPs in general. Analysis of RNA associated with polysomes in the absence of one of the two isoforms shows that the absence of the more abundant Tif4631p reduces polysome levels but there is no difference in proportion of the association of any mRNA between the two strains. Implying that there is no difference in function between these two paralogues in translation (Clarkson et al., 2010). It is possible that Tif4632p acts later in splicing, with U6, and so could have a different role in splicing and nuclear decay.

5.4.1 Future work

As indicated in the previous section, eIF4G's role in splicing needs to be confirmed; is splicing affected or degradation? One of the major barriers to separating the functions is the weak effect of deletion because the two can compensate for each other. Deletion of one, followed by β-est AID depletion (chapter 6), of the other would lead to much stronger defects. Splicing

monitored by any of the techniques in this thesis (reporter, microarray, RNA-seq, RT-qPCR of splicing intermediates and thio-labelling for kinetic evaluation), would clearly indicate which process(es) eIF4G acts on, and also if the two proteins have different areas of competence.

Another interesting field of study would be to see if the interaction sites in Figure 2 and Table 1 are mutually exclusive and, if they are, what role this interacting protein swap plays in RNA generation, processing and translation. Of interest to splicing is the potential Prp11p/RNA and Snu71p/CBC exchanges.

5.5 Contribution

I performed and analysed all the microarray experiments and primer extensions, P.Figure 6 and P.Table 2. The *in vitro* splicing reactions, P.Figure 5, were a collaboration between myself and Dr Kafasla.

All additional data presented in this chapter are my own, as is any re-interpretation of the results and conclusions from the publication.

5.6 References

Most references are in the publications, a few additional ones are listed here.

Berset, C., Zurbriggen, A., Djafazadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA *9*, 871–880.

Brameier, M., Krings, A., and MacCallum, R.M. (2007). NucPred—Predicting nuclear localization of proteins. Bioinformatics *23*, 1159–1160.

Clarkson, B.K., Gilbert, W.V., and Doudna, J.A. (2010). Functional Overlap between eIF4G Isoforms in Saccharomyces cerevisiae. PLOS ONE *5*, e9114.

Das, S., and Das, B. (2016). eIF4G—an integrator of mRNA metabolism? FEMS Yeast Res. *16* fow287.

Das, S., Saha, U., and Das, B. (2014). Cbc2p, Upf3p and eIF4G are components of the DRN (Degradation of mRNA in the Nucleus) in Saccharomyces cerevisiae. FEMS Yeast Res. *14*, 922–932.

Goyer, C., Altmann, M., Lee, H.S., Blanc, A., Deshmukh, M., Woolford, J.L., Trachsel, H., and Sonenberg, N. (1993). TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein

complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. Mol. Cell. Biol. *13*, 4860–4874.

Grifo, J.A., Tahara, S.M., Morgan, M.A., Shatkin, A.J., and Merrick, W.C. (1983). New initiation factor activity required for globin mRNA translation. J. Biol. Chem. *258*, 5804–5810.

Ohrt, T., Prior, M., Dannenberg, J., Odenwälder, P., Dybkov, O., Rasche, N., Schmitzová, J., Gregor, I., Fabrizio, P., Enderlein, J., et al. (2012). Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. RNA *18*, 1244–1256.

Park, E.-H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. EMBO J. *30*, 302–316.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl.Acids.Res. *22*, 4673–4680.

Will, C.L., and Lührmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. *3* a003707.

5.7 Reprint

RESEARCH PAPER

RNA Biology 6:5, 563-574; November/December 2009; © 2009 Landes Bioscience

Interaction of yeast eIF4G with spliceosome components

Implications in pre-mRNA processing events

Panagiota Kafasla,[1,3,*] J. David Barrass,[1] Elizabeth Thompson,[1] Micheline Fromont-Racine,[2] Alain Jacquier[2] Jean D. Beggs[1] and Joe Lewis[1,†]

[1]Wellcome Trust Centre for Cell Biology; University of Edinburgh; Edinburgh, UK; [2]Genetique des Interactions Macromoleculaires; Institut Pasteur (CNRS-URA2171); Paris, France; [3]Department of Biochemistry; University of Cambridge; Cambridge, UK

[†]Current address: Chemical Biology Core Facility; EMBL; Heidelberg, Germany

Key words: pre-mRNA processing, spliceosome, eIF4G, Prp11p, Snu71p

As evidenced from mammalian cells the eukaryotic translation initiation factor eIF4G has a putative role in nuclear RNA metabolism. Here we investigate whether this role is conserved in the yeast *Saccharomyces cerevisiae*. Using a combination of in vitro and in vivo methods, we show that, similar to mammalian eIF4G, yeast eIF4G homologues, Tif4631p and Tif4632p, are present both in the nucleus and the cytoplasm. We show that both eIF4G proteins interact efficiently in vitro with UsnRNP components of the splicing machinery. More specifically, Tif4631p and Tif4632p interact efficiently with U1 snRNA in vitro. In addition, Tif4631p and Tif4632p associate with protein components of the splicing machinery, namely Snu71p and Prp11p. To further delineate these interactions, we map the regions of Tif4631p and Tif4632p that are important for the interaction with Prp11p and Snu71p and we show that addition of these regions to splicing reactions in vitro has a dominant inhibitory effect. The observed interactions implicate eIF4G in aspects of pre-mRNA processing. In support of this hypothesis, deletion of one of the eIF4G isoforms results in accumulation of un-spliced precursors for a number of endogenous genes, in vivo. In conclusion these observations are suggestive of the involvement of yeast eIF4G in pre-mRNA metabolism.

Introduction

The cap (m^7GpppN) structure added co-transcriptionally to RNA polymerase II transcripts has been shown to influence many aspects of RNA metabolism, including pre-mRNA splicing,[1,2] 3' end formation,[3] export from the nucleus,[4-6] stability[7] and translation.[8] In the nucleus, the cap structure interacts with the predominantly nuclear cap-binding complex (CBC), a heterodimer consisting of cap-binding proteins CBP20 (Mud13p in *S. cerevisiae*) and CBP80 (Sto1p or Gcr3p in *S. cerevisiae*).[9,10] CBP20 is highly conserved from yeast to human, whereas CBP80 is far less conserved. CBP20 recognizes and binds capped RNA in conjunction with CBP80.[11] CBC plays a direct role in precursor messenger RNA (pre-mRNA) splicing, promoting the association of U1 small nuclear ribonucleoprotein particle (snRNP) with the cap-proximal 5' splice site.[2,12] In *Saccharomyces cerevisiae*, CBC interacts with Snu56p, a yeast-specific component of the U1 snRNP and a *cbp20*-Δ *cbp80*-Δ double mutant strain shows synthetic lethality with *SNU71*, a gene encoding for another component of the yeast U1 snRNP, Snu71p.[13] Furthermore, CBC exits from the nucleus to the cytoplasm together with the mRNA,[4-6] where it is thought to be replaced by the eIF4F complex.

In the cytoplasm, the effect of the cap structure on mRNA translation is mediated by a trimeric complex termed eukaryotic translation initiation factor 4F, eIF4F. eIF4F consists of the cap-binding subunit eIF4E, eIF4G and the RNA-helicase eIF4A.[14] eIF4G acts as a bridge between the cap structure and components of the ribosomal initiation complex.[14,15] In addition to eIF4E and eIF4A, eIF4G interacts with the poly(A)-binding protein PABP (Pab1p in yeast) facilitating the functional association of the 3' end of an mRNA with its 5' end to promote translation,[16] while the association between eIF4G and eIF4E markedly enhances the binding of the latter to the cap structure.[14] eIF4G and CBP80 are both characterized by the presence of the MIF4G domain, a structural motif known to be present in many proteins involved in RNA metabolism.[17,18] In *S. cerevisiae* there are two functionally redundant in translation isoforms of eIF4G, encoded by the genes *TIF4631* and *TIF4632*.[19]

A stable association of eIF4G with CBC detected in the nucleus of human cells plays possibly a role in coupling RNA-processing events in the nucleus with mRNA translation in the cytoplasm.[20] In mammalian cells, the interaction between eIF4G and CBC is required for the "pioneering" round of translation that leads to nonsense mediated decay (NMD). NMD is a

*Correspondence to: Panagiota Kafasla; Email: pk303@cam.ac.uk
Submitted: 06/01/09; Revised: 08/10/09; Accepted: 08/18/09
Previously published online: www.landesbioscience.com/journals/rnabiology/article/9861

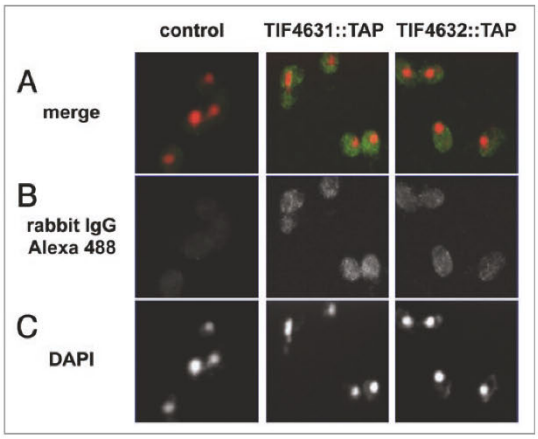


Figure 1. Yeast eIF4G homologues are distributed between the nucleus and the cytoplasm. Wild-type yeast cells (control) and cells expressing TIF4631:TAP or TIF4632:TAP were grown to OD_{600}: 0.2–0.4 and immobilized on slides as described in Materials and Methods. The location of Tif4631p and Tif4632p was detected with rabbit IgG Alexa 488 labeled (B). DAPI staining was used for the localization of the nuclei (C). The merged image is also presented in (A).

surveillance mechanism comprising the recognition and subsequent degradation of mRNAs bearing a premature termination codon.[21] Research in human cell lines suggests that both nuclear and cytoplasmic NMD occur on CBC-associated rather than eIF4E-associated mRNA, suggesting a role for CBC in translation.[22,23] In mammalian cells, NMD is translation and splicing dependent.[24-27] Ferraiulo et al.[28] reported that translation initiation factor eIF4AIII, a mammalian nucleo-cytoplasmic shuttling protein that interacts physically or functionally with eIF4G, is loaded onto the mRNA during splicing in the nucleus and then functions during NMD, indicating one more link between nuclear and cytoplasmic RNA processing events.

In *Saccharomyces cerevisiae* the domain of eIF4G responsible for the interaction with CBC resides between the eIF4E binding motif and the MIF4G domain.[29] Fortes et al.[29] proposed a role for the CBC-eIF4G interaction in the exchange of CBC for eIF4E and/or the direct recruitment of nascent mRNA for translation. However, it is not known whether this interaction occurs in the cytoplasm or in the nucleus. The latter can not be excluded since CBC as well as many components of the translation machinery are present in both cellular compartments.[4,30-32] In *S. cerevisiae* the interaction between eIF4G and CBC is not required for the first round of mRNA translation that proceeds NMD, while the exact role of this interaction is yet to be defined.[33] Yeast two-hybrid experiments have previously suggested that *S. cerevisiae* splicing factors, such as Prp11p and Snu71p could interact with the yeast eIF4G protein,[34] and the interaction of Prp11p with Tif4631p was proposed by Ho et al.[35] in a high throughput mass spectrometric protein complex identification screen. In human cells, association of eIF4G with pre-mRNA and the spliceosome, as well as partial co-localization of nuclear eIF4G with spliceosomal snRNPs are suggestive of a role for eIF4G in nuclear RNA-processing, perhaps in coupling splicing to other nuclear events and to translation.[20]

In the present work we assess the role of yeast eIF4G proteins in processing of RNA in the nucleus. We show the presence of eIF4G in the yeast nucleus and identify nuclear components that interact with eIF4G. We characterize in detail the interaction of eIF4G with protein and RNA components of the yeast spliceosome. We further investigate the possible role of yeast eIF4G in pre-mRNA splicing in vitro, and we show that depletion of one of the eIF4G homologues in vivo results in accumulation of intron containing pre-mRNAs for a number of endogenous genes. Our results suggest that yeast eIF4G has a role in pre-mRNA processing in the nucleus.

Results

Subcellular localization of Tif4631p and Tif4632p. To determine whether the yeast homologues of eIF4G, Tif4631p and Tif4632p, are present in both the cytoplasm and the nucleus, similarly to their human homologues, the proteins were expressed from their native promoter as C-terminally TAP-tagged fusion proteins and their localization was determined by indirect fluorescence and confocal microscopy. Both Tif4631p-TAP and Tif4632p-TAP were, as expected, found to be abundant in the cytoplasm (**Fig. 1A and B**). In addition, TAP-tagged Tif4631p and to a lesser extent Tif4632p could be both detected also in the nucleus, as shown in **Figure 1B**, in comparison to **C** where the location of the nuclei is indicated, and in the merged image **A**. This finding prompted us to determine whether the yeast eIF4G homologues can interact with components of the splicing machinery, similar to the situation observed in human extracts.[20]

Interaction of Tif4631p and Tif4632p with spliceosomal snRNPs. Whole cell extracts were made from strains expressing either Tif4631p-TAP or Tif4632p-TAP and the tagged proteins were precipitated using protein A Sepharose. An isogenic wild type strain was used as a negative control. Co-precipitated RNAs were purified, resolved on a denaturing gel, analysed by northern blot analysis with oligonucleotide probes specific to U1, U2, U4, U5 and U6 snRNAs (**Fig. 2A**) and the efficiency of precipitation was quantified (see legend of **Fig. 2B**). The eIF4G homologues reproducibly pulled down U1 snRNA. More specifically, ~20% of the input levels of U1 snRNA were pulled down by Tif4631p-TAP, whereas ~6% was precipitated by Tif4632-TAP (**Fig. 2A and B**), while more than 60% of the input U1 snRNA levels were precipitated under the same conditions by the TAP-tagged yeast CBP20 homologue, Mud13p (data not shown). The interaction of Tif4631p with U1 snRNA was reduced to ~10% of the input levels in increased salt concentration, whereas the less efficient interaction of this particular UsnRNA with Tif4632p could withstand salt much better, remaining at similar levels at 350 mM NaCl (**Fig. 2A and B**). Furthermore, ~7% of the input levels of U6 snRNA and ~10% of U4 snRNA were also pulled down by both eIF4G homologues, however these amounts were significantly reduced by increased salt concentration, mainly for Tif4631p and to a lesser extent for Tif4632p (**Fig. 2B**). Finally,

less than 5% of the input levels of U5 and U2 snRNAs were precipitated by the eIF4G homologues, and these interactions were greatly abolished when salt was increased to 350 mM NaCl (**Fig. 2B**). There was no significant precipitation of UsnRNAs from whole cell extracts prepared from the isogenic wild type strain (**Fig. 2A**, lanes 2 and 3) indicating that the detected RNAs were pulled down via their interaction with Tif4631-TAP and Tif4632-TAP proteins. These results suggest a specific interaction of both Tif4631p and Tif4632p mainly with U1 snRNA, with the interaction of Tif4631p and U1snRNA being more prominent (**Fig. 2**). In addition, both proteins can pull down to a similar extent U4 and U6 snRNAs (**Fig. 2B**).

Interaction of Tif4631p and Tif4632p with protein components of U1 and U2 snRNPs. There is evidence from yeast 2-hybrid screens that Tif4631p can interact specifically with the U2 snRNP protein, Prp11p and the U1 snRNP specific protein Snu71p, whereas Tif4632p was found to interact with Prp11p.[34] To assess these interactions GST-Prp11p and GST-Snu71p fusion proteins were expressed in *E. coli* and bound to Glutathione-Sepharose beads. The immobilized proteins were incubated with ^{35}S-labeled Tif4631p or Tif4632p and the purified complexes were analysed by SDS-PAGE and fluorography. Both Tif4631p and Tif4632p were pulled down by the immobilized GST-Prp11p very efficiently (**Fig. 3A**, lanes 1 and 2). Both proteins were also pulled down by GST-Snu71p, although in substantially lower amounts than with GST-Prp11p (**Fig. 3A**, lanes 7 and 8), whereas no protein was bound to Glutathione-Sepharose beads alone (lanes 4–6), indicating that the observed binding was specific. Incubation of the immobilized GST-Prp11p with ^{35}S labelled Snu71p and analysis of the purified complexes gave no detectable signal (**Fig. 3A**, lane 3), showing that the interaction detected between Tif4631p, Tif4632p and GST-Prp11p and -Snu71p is specific. We propose that there is a specific association between Prp11p and Tif4631p, characterized by the presence of two pools of complexes, since a great percentage of the interaction was lost at increased salt concentration

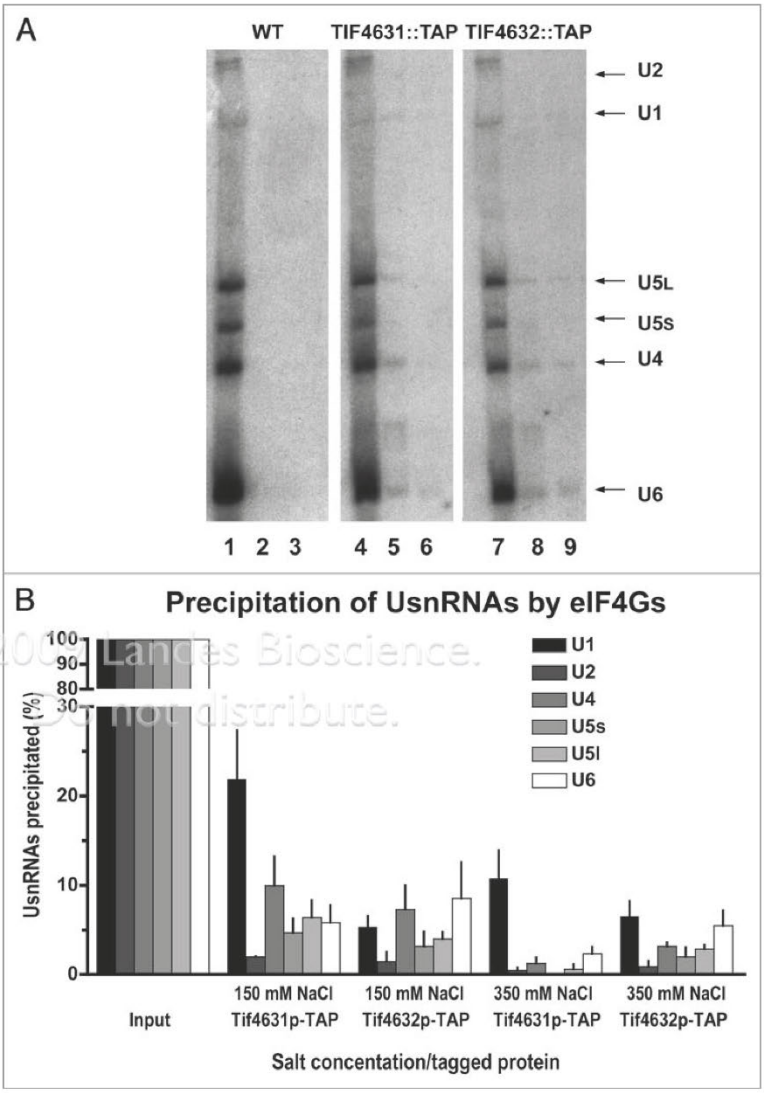


Figure 2. Yeast eIF4G homologues can precipitate U1 and U2 snRNA in vitro. (A) Yeast extracts derived from strains expressing either Tif4631p-TAP (lanes 4–6) or Tif4632p-TAP (lanes 7–9), as well as the isogenic wild-type strain (lanes 1–3), were incubated with IgG-Sepharose beads under increasing salt concentration (lanes 2, 5, 8: 150 mM NaCl, lanes 3, 6, 9: 350 mM NaCl). The co-precipitated RNAs, as well as 20% of the input RNAs (lanes 1, 4, 7), were assayed by northern blotting, using specific oligonucleotide-probes for each of the U1, U2, U4, U5 and U6 snRNAs. (B) Densitometry and Phosphorimager analysis were both used to quantify the signal produced by the northern analysis presented in (A), using the TotaLab software (Nonlinear Dynamics, UK) and ImageQuant software respectively. The results presented are means ± SEM from three independent experiments.

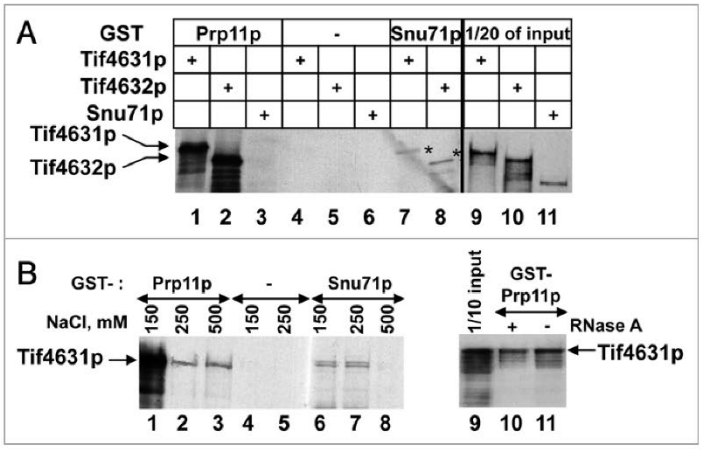


Figure 3. Yeast eIF4G homologues interact with protein components of U1 and U2 snRNPs. (A) Immobilized GST-Prp11p (lanes 1–3) or GST-Snu71p (lanes 7 and 8) were incubated with ^{35}S-labeled Tif4631p, Tif4632p or Snu71p as described in Materials and Methods. The pulled-down proteins were resolved by SDS-PAGE, and visualized by autoradiography. GST alone immobilized on Glutathione-Sepharose beads was used as negative control (lanes 4–6). 1/20 of input ^{35}S-labeled proteins were also analysed (lanes 9–11). (B) Immobilized GST-Prp11p (lanes 1–3, 10 and 11) or GST-Snu71p (lanes 6–8) or GST alone (lanes 4 and 5) were incubated with ^{35}S-labeled Tif4631p, in the presence of increasing amounts of NaCl, or in the presence (+) or absence (-) of 200 ng/ml RNase A as indicated. The pulled-down proteins were resolved by SDS-PAGE, and visualized by autoradiography. 1/10 of the input ^{35}S-labeled proteins were also analysed (lane 9).

(up to 500 mM NaCl, **Fig. 3B**, lanes 1–3) while a small but significant percentage was persistent even at very high salt concentration. The interaction of Tif4631p with Snu71p was less resistant to the highest salt concentration used (**Fig. 3B**, lanes 6–8). RNase A treatment did not significantly alter the binding pattern, demonstrating that the interaction of Tif4631p and Tif4632p with both Prp11p and Snu71p is not RNA mediated (**Fig. 3B**, lanes 9–11, and data not shown).

Mapping of the domain of Tif4631p that interacts with splicing factors. To map the domains of Tif4631p that are responsible for the interactions with both Snu71p and Prp11p, N-terminal and C-terminal truncated forms of Tif4631p were generated, in vitro translated and used in GST pull down experiments using immobilized GST-Prp11p or GST-Snu71p (**Fig. 4A**). Deletion of residues 657–952 from the C-terminus of Tif4631p had no effect on binding efficiency of this protein to either GST-Prp11p or GST-Snu71p. Further deletion of amino acids 593–656 resulted in about 30% less efficient binding, whereas complete loss of pull-down efficiency was evidenced when amino acids 267–592 were also deleted (**Fig. 4A**). In addition, deleting the N-terminal residues up to amino acid 452 did not affect the binding efficiency of Tif4631p for either Prp11p or Snu71p (**Fig. 4A**). Further deletion of amino acids 453–493 resulted in approximately 50% reduction of the interaction efficiency with Prp11p, while still maintaining the interaction with Snu71p (**Fig. 4A**). After sequential deletion of amino acids 494–528 the truncated Tif4631p could not be pulled down by GST-Prp11p, whereas deletions extending beyond amino acid 567 resulted in loss of interaction with GST-Snu71p (**Fig. 4A**). The above suggest that the minimal region of Tif4631p required for interaction with Prp11p resides between residues 494 and 529 (**Fig. 4A**).

To define in further detail the minimal domain of Tif4631p required for interaction with Prp11p, 10aa deletion mutants of Tif4631p were expressed and assayed in pull down assays with GST-Prp11p as described above. **Figure 4B** shows that individual deletion of aa 491–501, 534–544 or 566–576 reduced significantly the amount of Tif4631p mutant that could be pulled down by GST-Prp11p (lanes 2–4, compared to lane 1), suggesting that these residues contribute to the interaction between these proteins. It is noteworthy that none of these deletions was enough to abolish the interaction completely. In addition, substitution of residues 496–498 or 506–508 with Ala residues showed that these specific amino acids, and more crucially residues 506–508 are important for the interaction of Tif4631p with GST-Prp11p (**Fig. 4B**, lanes 6, 7 compared to lane 1). On the other hand, deletion of residues 457–467, 469–479 (**Fig. 4C**, lane 5 and data not shown), as well as 565–647 (**Fig. 4C**) did not affect significantly the levels of Tif4631p mutant that could be pulled down by GST-Prp11p, as expected (**Fig. 4A**). Taken together these findings suggest that the absolutely essential region of Tif4631p required for the interaction with Prp11p resides between amino acids 508–529. It is obvious, however, that all amino acids within the region of 494–529 contribute to the high affinity interaction of Tif4631p with Prp11p, since deletion of individual domains within this region, in the context of an otherwise full-length protein do not abolish the interaction completely. The fact that a Tif4631p mutant lacking aa 504 to 952 can not be pulled down by GST-Prp11p (**Fig. 4C**) and neither can be a mutant lacking residues 1–529 (**Fig. 4A**), verifies that residues 504–529 are absolutely required for the interaction of Tif4631p with GST-Prp11p in our pull down assays. Using the same approach we mapped the region of Tif4631p interacting with Snu71p between amino acids 567 and 647 (**Fig. 4A**) and that was verified by the finding that deletion of residues 565–647 of Tif4631p resulted in loss of the interaction of this protein with GST-Snu71p in our pull-down assays (**Fig. 4D**, lane 4).

Fortes et al.[29] found that amino acids 490–592 are required for the interaction of Tif4631p with yeast CBC. The fact that this region includes also the residues necessary for interaction with Prp11p and Snu71p, as described above, prompted us to investigate whether we could characterize in more detail the individual interactions of this domain. Using our Tif4631p truncated and deletion mutants we found that deletion of amino acids 491–501 resulted in more than 30% reduction of the interaction

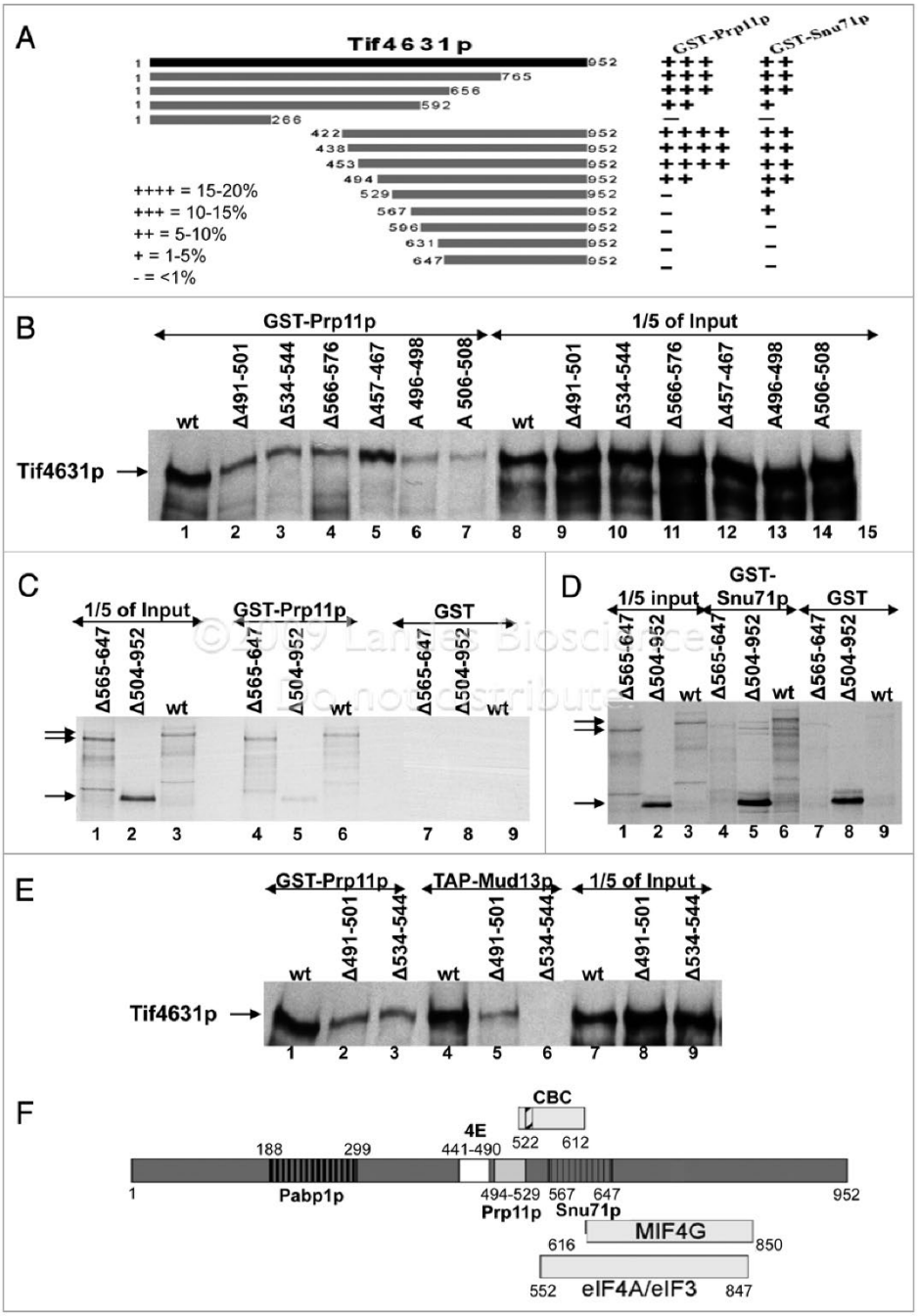


Figure 4. For figure legend, see page 568.

Figure 4. Detailed mapping of the interaction of Tif4631p with Snu71p and Prp11p. (A) Schematic representation of the deletion mutants of Tif4631p. Pull-down experiments were performed by incubation of immobilized GST-Pp11p or GST-Snu71p with ^{35}S-labeled full-length or truncated Tif4631p mutants and analysed as described in Figure 3A. Densitometry was used to quantify the respective autoradiographs and the quantification results are presented as % percentages of the input protein used for the pull-down experiments. (B) Immobilized GST-Prp11p (lanes 1–7) was incubated with ^{35}S-labeled Tif4631p (wt) or the Tif4632p-mutant proteins indicated, as described in Figure 3. The pulled-down proteins were resolved by SDS-PAGE, and visualized by autoradiography. 1/5 of input ^{35}S-labeled proteins were also analysed (lanes 8–15). (C) Immobilized GST-Prp11p (lanes 4–6) was incubated with ^{35}S-labeled Tif4631p (wt) or the Tif4632p mutants indicated. The pulled-down proteins were resolved by SDS-PAGE, and visualized by autoradiography. GST alone immobilized on Glutathione-Sepharose beads was used as negative control (lanes 7–9). 1/5 of input ^{35}S-labeled proteins were also analysed (lanes 1–3). The full length ^{35}S-labeled proteins are indicated by arrows. (D) Immobilized GST-Snu71p (lanes 4–6) were incubated with ^{35}S-labeled Tif4631p (wt) or the Tif4632p mutants indicated. The pulled-down proteins were resolved by SDS-PAGE, and visualized by autoradiography. GST alone immobilized on Glutathione-Sepharose beads was used as negative control (lanes 7–9). 1/5 of input ^{35}S-labeled proteins were also analysed (lanes 1–3). The full length ^{35}S-labeled proteins are indicated by arrows. (E) Immobilized GST-Prp11p (lanes 1–3) or TAP-Mud13p (lanes 4–6) were incubated with ^{35}S-labeled wild-type Tif4631p (wt) or Tif4631p mutants lacking amino acid residues 491–501 (Δ491-501), or 534–544 (Δ534-544). The pulled-down proteins were resolved by SDS-PAGE and visualized by autoradiography. 1/5 of the input ^{35}S-labeled proteins was also analysed for comparison. (F) Schematic representation of the domain of Tif4631p required for interaction with Prp11p, Snu71p and CBC. Other characterized domains of Tif4631p are also indicated (see text for details).

efficiency with TAP-Mud13p (**Fig. 4E**, lanes 4 and 5) and we identified the minimal region of Tif4631p required for interaction with TAP-tagged yeast CBP20 (Mud13p) in vitro to reside between residues 522 and 612 (data not shown). In addition, we identified a Tif4631p deletion mutant lacking amino acids 534–544, that could still bind to GST-Prp11p, but had abolished completely the interaction with TAP-Mud13p (**Fig. 4E**, lanes 3 and 6). This finding indicates that residues 534–544 of Tif4631p are essential for the interaction with yeast CBC. More importantly, the fact that Tif4631p-Δ(534-544) interacts with Prp11p but not with Mud13p shows that Tif4631p can interact with Prp11p in a CBC-independent manner. A summary of the above described domains, as well as a schematic representation of the domains of Tif4631p that are known to be required for interaction with Pabp1p[36] and eIF4E,[37] as well as the MIF4G domain[17] are shown for comparison in **Figure 4E**.

Sequence comparison shows that the region of Tif4631p that interacts with Prp11p, Snu71p and Mud13p is also very highly conserved in Tif4632p.[19] Using the same method as for Tif4631p, N-terminal and C-terminal truncation mutants of Tif4632p were assayed for their efficiency to bind to Prp11p and Snu71p and amino acids 424–609 were defined as the minimum region of Tif4632p required for interaction with Prp11p and Snu71p (data not shown).

Inhibition of splicing in vitro by Tif4631p/Tif4632p interaction domains. As shown in **Figure 3C**, the binding of $Tif4631_{494-952}$ to Prp11p was noticeably less than that observed with $Tif4631_{453-952}$. To obtain the minimum domain of Tif4631p that interacts efficiently with U1 and U2 snRNP proteins we constructed a Tif4631p domain mutant spanning amino acids 453 to 647, so as to include the region of Tif4631p that interacts with Snu71p. We refer to this construct as "Tif4631p(453-647)". Similarly, a construct with the Tif4632p domain required for interaction with both Prp11p and Snu71p was made and we refer to this construct as "Tif4632p(424-609)". After verifying, by pull-down experiments with GST-Prp11p and GST-Snu71 proteins, that domains Tif4631p(453-647) and Tif4632p(424-609) interact specifically with both Prp11p and Snu71p in vitro (data not shown), we used these constructs to determine whether Tif4631p(453-647) and Tif4632p(424-609) can influence splicing by supplementing in vitro splicing reactions with increasing amounts of these conserved domains fused to GST at their N-terminus. Addition of either GST-Tif4631p(453-647) (**Fig. 5A**, lanes 3–5) or GST-Tif4632p(424-609) (lanes 6–8) resulted in inhibition of splicing, as judged by the production of less intermediate (intron-3' exon) and final (intron-lariat) products of splicing, compared to the control (**Fig. 5A**, lane 2). Addition of GST alone had a minor effect in the efficiency of the splicing reaction (**Fig. 5A**, lanes 9–11). Consequently, addition of increasing amounts of either the Tif4631p or the Tif4632p domains in the splicing reaction resulted in an increasing inhibitory effect (**Fig. 5B**). Addition of the GST-Tif4631p mutant lacking amino acids 534–544, and therefore unable to interact with yeast CBP20, in in vitro splicing reaction showed that this dominant negative effect could not be attributed to selective sequestration of CBC by Tif4631p (data not shown). The above findings suggest that inhibition of splicing caused by Tif4631p(453-647) and Tif4632p(424-609) is possibly due to sequestration of spliceosomal components like Prp11p and Snu71p away from the spliceosomal machinery.

Deletion of TIF4631 results in accumulation of certain pre-mRNAs in vivo. The interaction of Tif4631p and Tif4632p with components of U1 and U2 snRNPs, together with the ability of Tif4631p(453-647) and Tif4632p(424-609) domains to inhibit splicing in vitro, suggested a possible role for the yeast eIF4G homologues in splicing. To test this hypothesis, given that double deletion of the yeast eIF4G homologues is lethal,[19] we assayed splicing in vitro using extract prepared from a strain deleted of TIF4631, and expressing TAP-tagged Tif4632p that could sequentially be depleted by IgG-Sepharose beads. Analysis of the RNA products of splicing showed no significant difference in the levels of splicing between control and TIF4631 deleted/Tif4632p depleted extracts, indicating that Tif4631p or Tif4632p could not be made limiting for efficient splicing in vitro. Western blot analysis performed to check the depletion level of Tif4632p in the extracts showed that we could not completely deplete Tif4632p. Consequently, we can not completely exclude the possibility that Tif4631p and Tif4632p play a role in splicing, since we possibly could not reduce yeast eIF4G homologues to limiting levels for splicing in vitro. We therefore focused our attention on splicing in vivo by assaying the splicing efficiency of different pre-mRNAs globally in the wild type and *tif4631*-Δ strains using

splicing microarrays.[38] Total RNA was extracted from the above strains and hybridized to slides containing an array of oligonucleotides able to hybridize to all the intron-containing RNAs of *S. cerevisiae* and to distinguish between pre-mRNAs and mRNAs.[38] These oligonucleotides were complementary to the 5' exon-intron junction of each one of the *S. cerevisiae* pre-mRNAs, or to the intron itself or to the exon-exon junction of each one of the mature mRNAs (**Fig. 6A**). For data normalization purposes oligonucleotides able to hybridize to the 3' exon of each one of all the pre-mRNAs and mRNAs were also used.

Out of 257 genes tested, there were 6 that consistently, in five independent experiments, showed accumulation of pre-mRNA at least two-fold or more in the *tif4631*-Δ strain compared to the isogenic control strain (**Table 2**). Similar analysis performed for a *tif4632*-Δ strain did not show any significant change in the pre-mRNA or mRNA levels, compared to the isogenic control strain. The microarray results were then verified for the gene YJR145C, for which accumulation of pre-mRNA could be detected in the *tif4631*-Δ strain compared to the wt strain, using primer extension experiments (**Fig. 6B**, lanes 1 and 2; quantified in **C**). As a positive control we used the *prp2-1* strain, that is deficient in splicing and exhibits a much stronger splicing block at the non-permissive temperature (37°C)[14] (**Figs. 6B**, lanes 3 and 4 and **5C**). YJR145C is the systematic name used for the gene *RPS4A*, which gives rise to the ribosomal protein Rps4Ap, a protein identical to Rps4Bp. The primer used for the primer extension experiment could hybridize to pre-mRNA and mRNA entities representing both genes. A two- to three-fold accumulation of the pre-mRNAs corresponding to both *RPS4A* and *RPS4B* could be detected in the *tif46321*-Δ strain compared to the isogenic control strain, whereas strain *prp2-1* showed the expected strong splicing deficient phenotype (**Fig. 6C**).

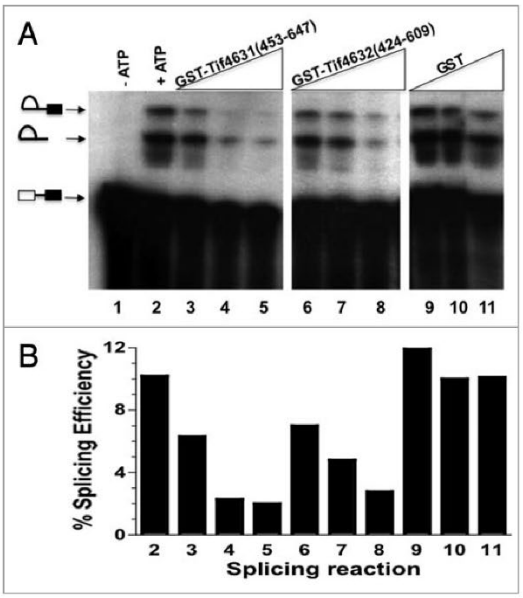


Figure 5. Tif4631p(453-647) and Tif4632p(424-609) inhibit splicing in vitro. (A) Increasing amounts (0.2–1 μg) of a GST-Tif4631p(453-647) (lanes 3–5) or GST-Tif4632p(424-609) (lanes 6–8) or GST alone (lanes 9–11), were added to in vitro splicing reactions, and the RNA products were analyzed by denaturing gel electrophoresis and autoradiography. Lane 1 shows the pre-mRNA used and lane 2 shows the control reaction without the addition of any protein. (B) Histogram derived from the quantification of data presented in (A) by phosphorimager analysis. Both spliced products (lariat and lariat exon) were quantified and are represented as a percentage of the unspliced pre-mRNA.

Discussion

In the present study, we investigate the putative role of the yeast translation initiation factor eIF4G in nuclear pre-mRNA processing. Using in situ localization we show that, in addition to their expected cytoplasmic localization, eIF4G homologues of *S. cerevisiae*, Tif4631p and to a smaller extent Tif4632p, exhibit a significant presence in the yeast nucleus. This agrees with previous reports that have demonstrated the presence of eIF4G in the nucleus of HeLa cells, indicative of a possible role for eIF4G in nuclear RNA processing.[20,39] Huh et al.[40] however, report that both yeast eIF4G homologues are located in the cytoplasm, using C-terminally GFP-tagged fusion constructs, fluorescent microscopy and an experimental approach different to ours.

In the cytoplasm, eIF4G plays an essential role in translation by acting as an adapter molecule during the initiation phase of protein synthesis. Within its sequence it contains domains that interact with eIF4E, eIF4A, eIF3, PABP and Mnk1.[14] In addition, *S. cerevisiae* eIF4G has been shown to interact with the cap binding complex, CBC, via a domain between eIF4E and eIF3 binding sites.[29] McKendrick et al.[20] have shown that, in mammalian cells, a nuclear pool of eIF4G is closely associated with CBC. Moreover, it is known that CBC in the nucleus promotes association of U1 snRNA with the cap proximal 5' splice site.[2,12]

Consistent with these results, we provide evidence here for the interaction of yeast eIF4G homologues with spliceosomal UsnRNPs. We show that mainly Tif4631p, and Tif4632p to a lesser extent, interact with U1 snRNA. In addition, both eIF4G homologues are also shown to interact stably with the U1 snRNP component Snu71p, either directly or via other proteins present in the reticulocyte lysate used in our assays. This particular interaction, however, is much less efficient possibly because Snu71p alone does not mediate the interaction of eIF4G proteins with U1 snRNP. Using pull-down binding assays, we show that both eIF4G homologues interact very efficiently and stably with Prp11p, a U2 snRNP component in an RNA-independent manner, whereas our TAP-tagged eIF4G homologues can only pull down a minor amount of U2 snRNA in vitro. This finding suggests that the high affinity interaction of yeast eIF4G proteins with Prp11p is independent of its interaction with U2 snRNA. The finding that eIF4G homologues can pull down U4 and U6 snRNAs to a similar extent is not unexpected since the existence of a penta-snRNP complex has been proposed for yeast.[41]

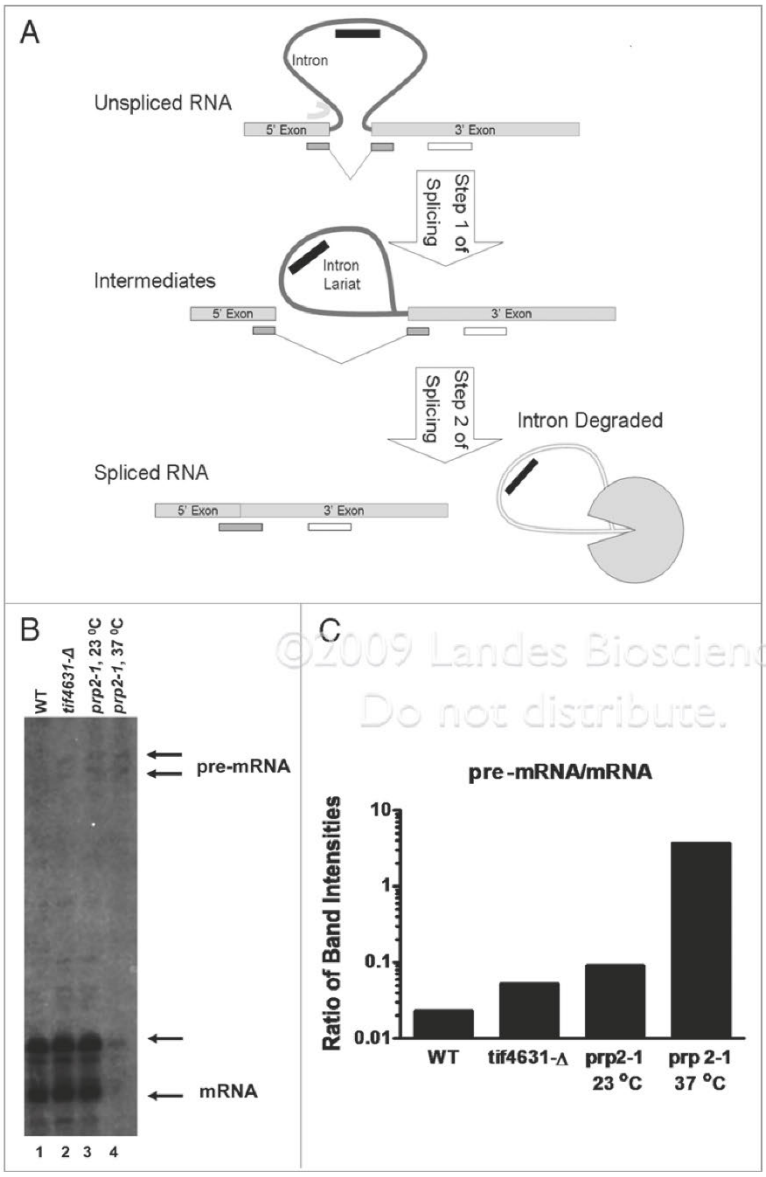


Figure 6. Accumulation of pre-mRNAs in *tif4631-Δ* strain in vivo. (A) Schematic overview of the splicing microarray. Four different probes able to detect intermediate species of the two stages of splicing were designed and used on the microarray. The intron probe (black), as well as the 5'-exon-intron probe (dark grey) can only be hybridized to by un-spliced RNA. The mature junction probe (grey) will only be hybridized to by processed mRNA, whereas the exon 2 probe (white) will be hybridized to by both pre-mRNA and mRNA and is used as a control to remove effects due to differential gene expression. (B) Total RNA was extracted from the *tif4631-Δ* strain as well as from the isogenic wild-type and was used in primer extension experiment, as described in Materials and Methods. The cDNAs were analysed by denaturing gel electrophoresis and autoradiography. RNA from strain *prp2-1*, grown either at 23°C or at 37°C was also used in the same experiment as a control. The positions of both the mature and precursor mRNAs for both *RPS4A* and *RPS4B* are indicated. (C) Quantification of the bands corresponding to both pre-mRNAs and mRNAs of the gel presented in (B) by phosphorimaging.

Fine mapping of the individual domains of eIF4G proteins required for the interaction with Prp11p, CBC and Snu71p shows that they are distinct, but adjacent to each other and together they form an "interaction domain" that resides in a region that is highly conserved between the two yeast eIF4G isoforms. Addition of a recombinant form of the above defined "interaction domain" of yeast eIF4G can efficiently inhibit splicing in vitro, possibly by interacting with spliceosomal components and sequestering them away from the splicing machinery, supporting the hypothesis that eIF4G proteins can participate in pre-mRNA processing events.

To further investigate the possible role of yeast eIF4G in splicing, we undertook a global analysis of pre-mRNA splicing using microarrays. A subset of six pre-mRNAs showed a reproducible two- to three-fold accumulation of pre-mRNAs in strains deleted of the *TIF4631* compared to the isogenic control (**Table 2**). Even though the possibility of this being an indirect effect can not be ruled out, it is interesting that the same result was not observed when this microarray analysis was applied on a strain deleted of the *TIF4632*, where no accumulation of any pre-mRNA was observed. Our results show that predominantly Tif4631p interacts with spliceosomal U1 snRNP in vitro, and this interaction could be possibly required for efficient processing of certain pre-mRNAs. Analysis and comparison of intron sequences and splice sites of the pre-mRNAs presented in **Table 2** did not reveal any obvious common features between these RNAs, apart from the fact that they are all pre-mRNAs that will give rise to ribosomal proteins. In addition,

five of these RNAs are characterized by the presence of a non-consensus 5' splice site. The interaction of eIF4G with UsnRNPs during splicing in vivo could be possibly necessary as a check point during the processing of these pre-mRNAs. In support of this, a partial but significant co-localization of eIF4G with a proportion of snRNPs at discrete foci was reported in mammalian cells,[20] a finding that strengthens the hypothesis that eIF4G might participate in the processing of some pre-mRNAs in the nucleus.

In conclusion, we present here a detailed mapping of the domains of yeast eIF4G required for interaction with components of the spliceosomal machinery. These interactions probably implicate eIF4G in nuclear RNA processing events in *S. cerevisiae*. Even though yeast eIF4G interacts specifically with spliceosomal components, it is not required for splicing of the RNAs tested in vitro, but a small but significant accumulation of pre-mRNAs is observed in a *tif4631*-Δ strain in vivo. A *tif4632*-Δ strain does not exhibit the same effect, a finding that might explain the normal growth phenotype of this strain, in contrast to the slow growth phenotype of *tif4631*-Δ strain reported by Goyer et al.[19] At present we can not explain why the genetic depletion of one of the eIF4G homologues in yeast affects the splicing of a small number of pre-mRNAs that correspond to ribosomal proteins. There is evidence of an auto-regulatory role for alternative splicing coupled NMD in the expression of a number of ribosomal proteins,[42-44] while MIF4G,[17] a domain that is conserved between eIF4G, Upf2p/NMD2p and CBP80 implicates eIF4G and CBP80 in nonsense-mediated decay. One could suggest that eIF4G participates in nuclear RNA processing by linking in some way pre-mRNA splicing with degradation pathways like NMD and acting once more as a scaffold protein that brings important protein components together.

Materials and Methods

Plasmids and *S. cerevisiae* strains. Plasmids pBS-Tif4631 and pBS-Tif4632 were used as templates for coupled in vitro transcription-translation reactions in rabbit reticulocyte lysates to express [^{35}S]-labeled eIF4G proteins. They were constructed by subcloning PCR amplified *TIF4631* and *TIF4632* genomic DNAs respectively into pBS(SK-) (Stratagene) as EcoR1/BamH1 fragments with primers 5'-GGG AAT TCA TGA CAG ACG AAA CTG TCA AC-3' and 5'-CCG GAT CCT TAC TCT TCG TCA TCA CTT TCT-3' for *TIF4631* and 5'-GGG AAT TCA TGA CTG ACC AAA GAG GTC CAC-3' and 5'-CCG GAT CCT TAA TCA CTG TCC CCA TCG TTA-3' for TIF4632. Plasmid pGEX4T-1-Prp11 was used to express GST-Prp11p. The ORF of PRP11 was subcloned as a Sal1/SmaI fragment from pAS2DD-prp11,[34] into pGEX4T-1 XhoI/SmaI restriction sites. To produce plasmids expressing C-terminally truncated versions of Tif4631, pBS-T1F4631 was digested with each of BstBI, StuI, NheI and BlpI to obtain the 1–765, 1–656, 1–592 and 1–266 truncated forms of Tif4631p respectively. To obtain N-terminally truncated versions, a reverse primer 5'-C*TT A*CT CTT CGT CAT CAC TTT CTC CC-3' and the following forward primers were used to amplify by PCR the fragments of TIF4631 referred next to each primer:

Table 1. *S. cerevisiae* strains used in this study

Name	Genotype
TIF4631::TAP	*MATa ade2-101 his3-Δ200 leu2-Δ1 trp1-Δ99 ura3-Δ 99 cir^{0}tif4631-TAP::TRP1*
TIF4632::TAP	*MATa ade2-101 his3-Δ200 leu2-Δ1 trp1-Δ99 ura3-Δ 99 cir^{0}tif4632-TAP::TRP1*
tif4631-Δ	*MATalpha ade2 his3 leu2 trp1 ura3(GAL$^+$) tif4631Δ::HIS3*
tif4632-Δ	*MATalpha ade2 his3 leu2 trp1 ura3(GAL$^+$) tif4632Δ::HIS3*
tif4631-Δ; TIF4632::TAP	*MATa ade2-101 his3-Δ200 leu2-Δ1 trp1-Δ99 ura3-Δ 99 cir^{0}tif4632-TAP::TRP1 tif4631Δ::HIS3*

5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* ACA CCA ATT GAA GAT GTC-3' for amino acids 422–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* GGT CCT GAT ATC AAA TAC for amino acids 438–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* CCA ACT TTC TTG CTT C-3' for amino acids 453–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* GGA GAT TCT GGC AGA TTC GGC-3' for amino acids 494–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* AGA AGA TCA AAG AGA-3' for amino acids 529–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* AAG GAA GAA GTT GCT CC-3' for amino acids 567–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* GAC GGA AAG ACC GAC TAT TGG-3' for amino acids 596–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* GCT ATT GCA AAC ATA TCA G-3' for amino acids 631–952, 5'-GGA TCC TAA TAC GAC TCA CTA TAG GAA CAG ACC ACC *ATG* GCT GTG ATA GAA CAG-3' for amino acids 647–952. The PCR products were purified and used directly in in vitro transcription/translation reactions with rabbit reticulocyte lysates according to manufacturer's instructions (Promega) to obtain the ^{35}S-labeled truncated versions of Tif4631p.

To construct deletion mutant Tif4631p-Δ(565-647), the PCR product produced by forward primer 5'-TAT ACA TAT GGC CCA ACT TTC-3' and reverse primer 5'-CGA CGA TAT CGT CCT CTC TTT TC-3' on pBS-Tif4631 as a template, was digested with NdeI/EcoRV and the PCR product produced by forward primer 5'-CAC CGA TAT CGA ACA GAT TTT C-3' and reverse primer 5'-CGC AGG ATC CAA GAG AAT GAA TGA C-3' was digested with EcoRV/SacII. The digested PCR products were ligated with each other and the pBS-Tif4631 fragment derived from NdeI/SacII digest. For construction of pBS-Tif4631-Δ(504-952), pBS-Tif4631 was digested with MscI/StuI and the respective fragment was gel extracted and ligated.

The strains of *S. cerevisiae* used in this study are described in **Table 1**. Standard yeast growth conditions and manipulations were used.

In situ localization. Yeast strains TIF4631::TAP or TIF463::TAP were grown to OD_{600} = 0.2–0.4 and paraformaldehyde (EMS) was added to final concentration 4%. Cells were harvested, washed in buffer B (1.2 M sorbitol, 65 mM KH_2PO_4, 35 mM K_2HPO_4) at 4°C and resuspended in 1 ml of buffer B.

Table 2. Genes showing a splicing deficiency in the yeast strain *tif4631*-Δ, their respective ORFs and the ratio of pre-mRNA in the *tif4631*-Δ strain compared to the wt strain

Gene	ORF	*tif4631*-Δ/wt	Exon 1	5' splice site	Branch point	3' splice site	Intron size
YBR181C	RPS6B	2.06	gaag	guaugua	uuuacuaacaa	guauuauuuauaacag	352
YOR096W	RPS7A	2.07	agaa	guauguu	cuuacuaacat	uuuccuucuuuuauag	401
YGL189C	RPS26A	2.22	agua	guauguu	guuacuaacua	augaauauuuaauuag	368
YLR448W	RPL6B	2.33	acaa	guaugug	uauacuaacua	gauaugucaauuauag	384
YHR203C	RPS4B	2.89	gacc	guauguu	uuuacuaacaa	acgauuuuucauauag	269
YJR145C	RPS4A	3.16	gacc	guauguu	uuuacuaacga	auuuuuuuccguacag	256

Sequences of the 5' splice site, the branch point and the 3' splice sites, as well as the size of the corresponding introns are shown (according to Lopez et al.[51]).

After incubation with 1 mg oxalyticase for 10 min at 30°C, cells were washed with ice-cold buffer B and resuspended in 0.65 ml of the same buffer. A 0.1 ml aliquot was put on to a coverslip, allowed to stand for 30 min and then washed with 3 ml of buffer B, replaced with 5 ml of methanol and left for 5 min at -20°C. After a final wash with buffer B, cells were permeabilized by buffer C (0.1% Triton X-100, 20 mM HEPES pH 7.9, 200 mM NaCl) for 30 min. Primary antibody (Rabbit IgG Alexa 488 labelled, Molecular Probes) was used in 1:100 dilution in buffer C, for 1 hr. Washes with buffer C followed and after briefly rinsing with PBS the coverslip was mounted with vectashield with DAPI and examined using a Leica confocal microscope with three lasers giving excitation lines at 380, 488 and 543 nm. The data from the channels were collected separately using narrow-bandpass filter settings. In multiple staining experiments, the laser intensities and data collection settings were adjusted to avoid overlap (bleedthrough) between channels. The coupled microscope was a Leica DMIBRE equipped with a 633 water immersion objective lens (numerical aperture, 1.4). Data sets were processed using the Leica TCS NT, version 1.4.338, software package and were subsequently exported into Adobe Photoshop version 5.5, and Deneba Canvas version 7.02.

In vitro pull down. For purification of GST fusion proteins, *Escherichia coli* extracts from BL21-codon plus cells expressing either GST-Prp11p or GST-Snu71p fusion proteins were prepared by inducing 1 liter culture of exponentially growing cells at OD_{600} of 0.6–0.8, for 3.5 hrs with 1 mM IPTG. Cells were harvested, resuspended in PBS with 0.5% Triton X-100 and lysed by French press. The lysate was incubated with 500 µl of Glutathione-Sepharose beads for 2 hrs at 4°C. The beads were washed extensively with PBS-0.5% Triton X-100 and used subsequently in the binding assays. For in vitro transcription/translation reactions of *TIF4631* or *TIF4632*, the TNT® system (Promega) and plasmids pBS-TIF4631 or pBS-TIF4632 were used according to manufacturer's instructions. For binding assays, GST-fusion proteins bound to glutathione-sepharose beads were incubated with Tif4631p or Tif4632p produced with the TNT® system and the reaction mixtures were incubated for 2 hrs at 4°C. After extensive washing with PBS-0.5% Triton X-100, the proteins were eluted by SDS sample buffer and analysed by SDS/PAGE. For the detection of ^{35}S proteins, the EA starter kit was used (Thistle Scientific).

Purification of Tif4631p-TAP, Tif4632p-TAP proteins. Yeast strains TIF4631::TAP and TIF4632::TAP were grown to OD_{600} = 2 and cells were harvested and lysed in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM PMSF, 1 µM leupeptin, 2 µM pepstatin A, 4 µM chymostatin, 2.6 µM aprotinin) by French press. The TAP-tagged proteins were purified by incubating the yeast cell lysates with 50 µl of a 50% solution of IgG Sepharose beads for 2.5 hrs at 4°C. To analyse the RNAs that are pulled down by Tif4631-TAP or Tif4632-TAP proteins, the beads were washed extensively with either IPP150 buffer (150 mM NaCl, 0.1% Nonidet P-40, 10 mM Tris, pH 8.0) or IPP350 buffer (350 mM NaCl, 0.1% Nonidet P-40, 10 mM Tris, pH 8.0) and then the RNA was extracted by phenol/chloroform, precipitated and analysed on a 6% acrylamide gel. Northern hybridization followed using the following specific oligonucleotide probes: 5'-CTT AAG GTA AGT AT-3' for U1, 5'-CTA CAC TTG ATC TAA GCC AAA AGG C-3' for U2, 5'-AAT ATG GCA AGC CC-3' for U5, 5'-CTC TTT GTA AAA CGG TTC-3' for U6 and 5'-CCG TGC ATA AGG AT-3' for U4snRNA.

Preparation of whole cell extracts of yeast. The protocol of Umen and Guthrie[45] was used with minor modifications. Namely, the yeast strain of interest was grown to OD_{600} = 0.5–1 at 30°C in 1L YPD medium. Cells were harvested by centrifugation at 5,000 rpm for 5 min (Beckman JLA10.500 rotor) and washed twice with 20 ml of AGK buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 10% Glycerol, 2 mM DTT). Cells were resuspended in 0.4 cell pellet volume of AGK buffer and squirted through a syringe into liquid nitrogen. The frozen cell pellet was ground to a fine powder by mortar and pestle and allowed to thaw. The thawed powder was centrifuged at 17,000 rpm for 30 min at 2°C (rotor JA25.50 Beckman). The supernatant was centrifuged at 40,000 rpm for 1 hr at 4°C (rotor 70.1Ti, Beckman) and subsequently dialysed for 3 hrs against buffer D (20 mM HEPES pH 7.0, 50 mM KCl, 0.2 mM EDTA, 20% (v/v) glycerol) at 4°C, centrifuged at 13,000 rpm for 10 min and frozen at -80°C.

In vitro splicing. In vitro splicing reactions were carried out and analysed as described by Lin et al.[46] Actin pre-mRNA was transcribed from linearized by BamHI plasmid p283,[47] which contains an AluI fragment of the yeast actin gene cloned at a SmaI site of pGEM vector.

Microarrays. Total RNA was prepared as described by Schmitt et al.[48] while the microarrays used were similar to those described by Clark et al.[49] More precisely, there were at least four

oligonucleotide probes per *S. cerevisiae* transcript, containing an intron, immobilized on to the microarray. One probe was complementary to the 5'exon-intron junction of all the pre-mRNAs, one probe was complementary to sequences of the intron, another probe was complementary to the exon-exon junction of all the mature mRNAs and finally there was also a probe that would hybridize to the 3' exon of all the pre-mRNAs and mRNAs used for normalization. For transcripts containing more than one intron and/or different predictions of intron start and end points, the 5' and mature exon-exon junction probes (and occasionally the intron probe) were duplicated as necessary to identify all possible introns. These probes were synthesized by MWG-Biotech AG and have an average length of 41.2 nucleotides. Ideally a probe should have a Tm of 80–90°C (nearest neighbor algorithm), where constraints of the target sequence would allow. Similarly all the probes to a transcript were designed to have a Tm within ±2°C of each other where possible. These probes were printed in triplicate on to poly-lysine slides (Sigma) in 3x SSC buffer at the GTI (Chancellor's Building, Royal Infirmary, Edinburgh) using a GMS 417 (Affymetrix) printer. This microarray was hybridized with Cy dye labeled cDNA derived from total RNA from the mutant strain of interest and an isogenic wild type. The reverse transcription reaction was performed in the presence of Cy dye modified dUTP (Amersham). Either Cy3 or Cy5 was used to label the mutant and the other to label the wild type cDNA, including at least one occasion where the dyes were swapped from their normal total RNA type (dye flip labeling). Thermoscript (Invitrogen) was used to perform the reverse transcription reaction at 47°C using the manufacturer's supplied buffers and 20–50 µg of total RNA. Priming was from gene specific primers designed to bind downstream to the 3' exon probe (45 µM total concentration). The RNA was removed using 55 mM NaOH, 28 mM EDTA pH 8 and 65°C for 30 minutes before neutralization with 120 mM Tris pH 7.5 at 4°C. The labeled cDNAs (mutant and wild type), were combined and spun in YM-100 columns (Millipore), using the manufacturer's instructions, in 10 mM Tris pH 7.5 (used to wash twice more) to filter-purify and concentrate. The arrays were blocked with 1-methyl-2-pyrrolidinone (93% v/v) and succinic anhydride (1.64% w/v) in 68.5 mM boric acid (pH 7 with NaOH) at room temperature for 20 minutes and then pre-hybridized by 5x SSC, 0.1% (w/v) SDS, 1% (w/v) BSA at 47°C for 45 minutes before hybridization. The hybridization solution was 5x SSC, 0.4% (w/v) SDS and 20 µg of human COT-1 DNA in 16 µl under a 22 x 22 mm coverslip at 47°C for 10–16 hrs in a humid chamber. Then the arrays were washed in 1x SSC, 0.2% SDS followed by 0.1x SSC, 0.2% SDS and finally 0.1x SSC, all at room temperature for 5 minutes. Scanning was performed in an Affymetrix 418 scanner and spot quantitation was carried out by ImageQuant v3.1.

To normalize, the background corrected intensity value for each spot was divided by the background corrected intensity value from the corresponding 3'exon probe. More precisely, in a mutant strain where the pre-mRNA of a specific gene is more than the respective mRNA, one would see more RNA hybridized to the oligonucleotides that complement the 5'exon-intron junction or/and intron rather than the one complementary to the exon-exon junction, when compared to the wt strain. The above signals were compared to each other after normalization by the signal the specific RNA gives with the oligonucleotide that hybridizes to each 3' exon and recognizes it in both pre- and mature forms. Additionally, the signal that this RNA gives on the position of the 3'exon recognizing oligonucleotide probe should be equal to the sum of the signal it gives with the 5'exon-intron recognizing oligonucleotide probe, plus the signal from the hybridization with the exon-exon junction recognizing oligonucleotide probe. On the other hand, the same amount of RNA is expected to hybridize to the oligonucleotide probe that complements the 3' exon of this RNA for both control and mutant strain, since the total amount of the specific RNA should be equal in both strains.

For the primer extension experiments performed to check the microarray results, total RNA prepared as described above was hybridized with a ^{32}P-labeled oligonucleotide probe with the sequence 5'-AGA AAG ACA ATC AAT GGC AAG GA-3' that hybridizes in exon 2 of the YJR145C pre-mRNA. The reverse transcription reaction was performed as described by Bousquet-Antonelli et al.[50]

Acknowledgements

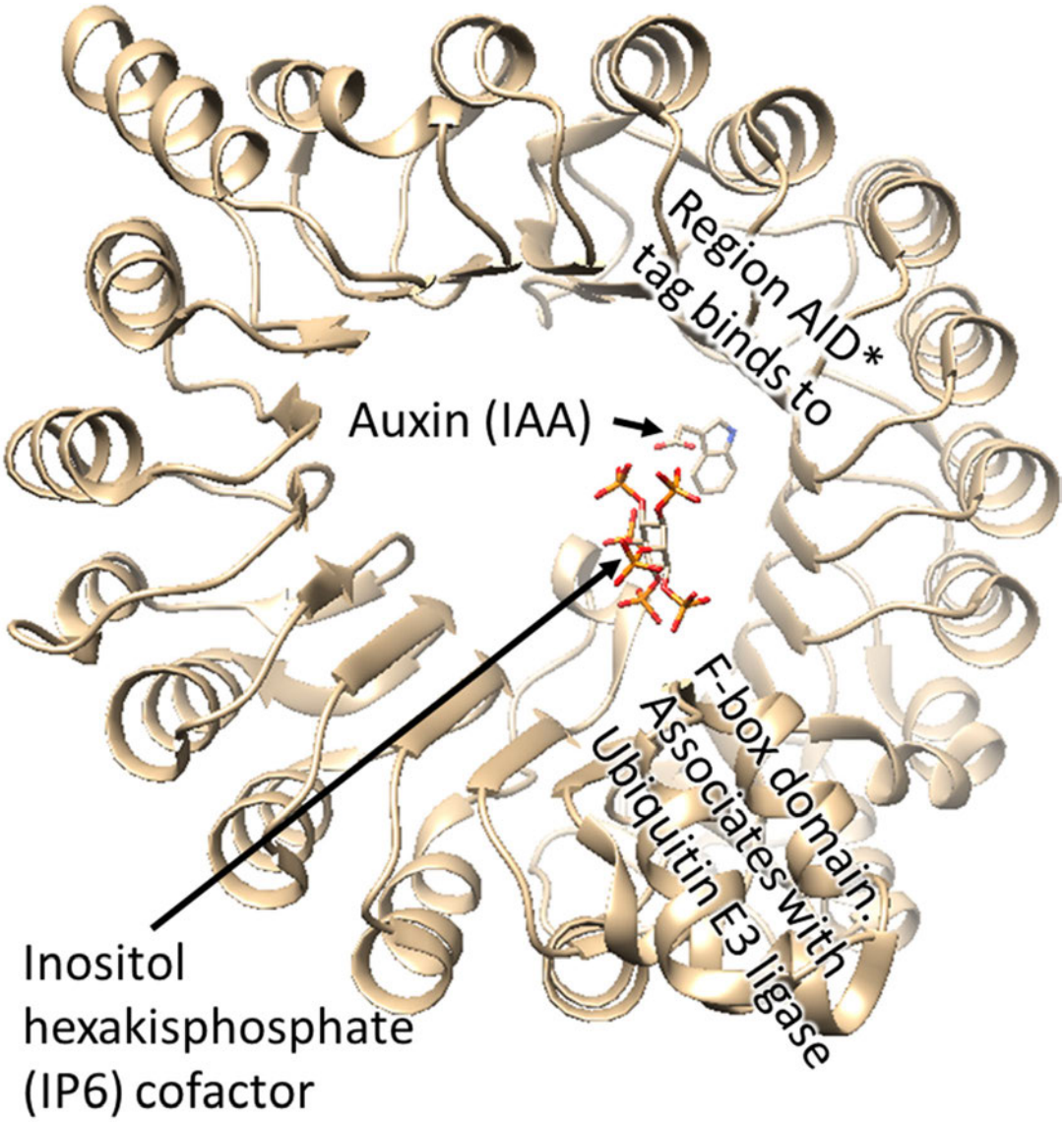
We thank Heike Roesner for help with the mapping of eIF4G interaction domains. We thank Skarlatos G. Dedos, and Tuija A.A. Pöyry for critical reading and comments on the manuscript. We also thank Prof. Richard J. Jackson (Dept. of Biochemistry, University of Cambridge) for his support. Research in the laboratory of J.L. was initially supported by the Medical Research Council. E.T. and A.J. were funded by the RNOMICS project QLG2-CT-2001-01554. J.D.B. was funded by the Wellcome Trust grant 067311 to Prof. Jean Beggs. P.K. held a Marie-Curie Fellowship (R81517).

References

1. Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 1994; 78:657-68.
2. Lewis JD, Gorlich D, Mattaj IW. A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucleic Acids Res 1996; 24:3332-6.
3. Flaherty SM, Fortes P, Izaurralde E, Mattaj IW, Gilmartin GM. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci USA 1997; 94:11893-8.
4. Gorlich D, Kraft R, Kostka S, Vogel F, Hartmann E, Laskey RA, et al. Importin provides a link between nuclear protein import and U snRNA export. Cell 1996; 87:21-32.
5. Visa N, Izaurralde E, Ferreira J, Daneholt B, Mattaj IW. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J Cell Biol 1996; 133:5-14.
6. Shen EC, Stage-Zimmermann T, Chui P, Silver PA. The yeast mRNA-binding protein Npl3p interacts with the cap-binding complex. J Biol Chem 2000; 275:23718-24.
7. Furuichi Y, LaFiandra A, Shatkin AJ. 5'-Terminal structure and mRNA stability. Nature 1977; 266:235-9.
8. Shatkin AJ. mRNA cap binding proteins: Essential factors for initiating translation. Cell 1985; 40:223-4.
9. Ohno M, Kataoka N, Shimura Y. A nuclear cap binding protein from HeLa cells. Nucleic Acids Res 1990; 18:6989-95.
10. Colot HV, Stutz F, Rosbash M. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes and Dev 1996; 10:1699-708.
11. Mazza C, Segref A, Mattaj IW, Cusack S. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J 2002; 21:5548-57.
12. Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev 1996; 10:1683-98.

13. Fortes P, Kufel J, Fornerod M, Polycarpou-Schwarz M, Lafontaine D, Tollervey D, Mattaj IW. Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol 1999; 19:6543-53.
14. Gingras AC, Raught B, Sonenberg N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem 1999; 68:913-63.
15. Dominguez D, Kislig E, Altmann M, Trachsel H. Structural and functional similarities between the central eukaryotic initiation factor (eIF)4A-binding domain of mammalian eIF4G and the eIF4A-binding domain of yeast eIF4G. Biochem J 2001; 355:223-30.
16. Hentze M. eIF4G: a multipurpose ribosome adapter. Science 1998; 275:500-1.
17. Ponting CP. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem Sci 2000; 25:423-6.
18. Aravind L, Koonin EV. Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. Genome Res 2000; 10:1172-84.
19. Goyer C, Altmann M, Lee HS, Blanc A, Deshmukh M, Woolford JL Jr, et al. TIF4631 and TIF4632: Two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. Mol Cell Biol 1993; 13:4860-74.
20. McKendrick L, Thompson E, Ferreira J, Morley SJ, Lewis JD. Interaction of eukaryotic translation initiation factor 4G with the nuclear cap-binding complex provides a link between nuclear and cytoplasmic functions of the m(7)guanosine cap. Mol Cell Biol 2001; 21:3632-41.
21. Lejeune F, Ranganathan AC, Maquat LE. eIF4G is required for the pioneer round of translation in mammalian cells. Nat Struct Mol Biol 2004; 11:992-1000.
22. Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell 2001; 106:607-17.
23. Lejeune F, Ishigaki E, Li X, Maquat LE. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: Dynamics of mRNP remodeling. EMBO J 2002; 21:3536-45.
24. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 1999; 96:307-10.
25. Maquat LE, Carmichael GG. Quality control of mRNA function. Cell 2001; 104:173-6.
26. Schell T, Kulozik AE, Hentze MW. Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. Genome Biol Reviews 2002; 3:1006.
27. Maquat LE. Nonsense-mediated mRNA decay in mammals. J Cell Science 2005; 118:1773-6.
28. Ferraiuolo MA, Lee C-S, Lian LW, Hsu JL, Costa-Mattioli M, Luo M-J, et al. A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. PNAS 2004; 101:4118-23.
29. Fortes P, Inada T, Preiss T, Hentze M, Mattaj IW, Sachs AB. The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Mol Cell 2000; 6:191-6.
30. Iborra FJ, Jackson DA, Cook PR. Coupled transcription and translation within nuclei of mammalian cells. Science 2001; 293:1139-42.
31. Dahlberg JE, Lund E, Goodwin EB. Nuclear translation: What is the evidence? RNA 2003; 9:1-8.
32. Nathanson L, Xia T, Deutscher MP. Nuclear protein synthesis: A reevaluation. RNA 2003; 9:9-13.
33. Baron-Benhamou J, Fortes P, Inada T, Preiss T, Hentze MW. The interaction of the cap-binding complex (CBC) with eIF4G is dispensable for translation in yeast. RNA 2003; 9:654-62.
34. Fromont-Racine M, Rain J-C, Legrain P. Towards a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genetics 1997; 16:277-82.
35. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002; 415:180-3.
36. Tarun SZ Jr, Sachs AB. Association of the yeast poly(A)-tail binding protein with translation initiation factor eIF-4G. EMBO J 1996; 15:7168-77.
37. Mader S, Lee H, Pause A, Sonenberg N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4gamma and the translational repressors 4E-binding proteins. Mol Cell Biol 1995; 15:4990-7.
38. Barrass JD, Beggs JD. Splicing goes global. Trends Genet 2003; 19:295-8.
39. Etchison D, Etchison JR. Monoclonal antibody-aided characterization of cellular p220 in uninfected and poliovirus-infected HeLa cells: subcellular distribution and identification of conformers. J Virol 1987; 61:2702-10.
40. Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. Global analysis of protein localization in budding yeast. Nature 2003; 425:686-91.
41. Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, Lee TD, Abelson J. Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell 2002; 9:31-44.
42. Cuccurese M, Russo G, Russo A, Pietropaolo C. Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. Nucleic Acids Res 2005; 33:5965-77.
43. Dabeva MD, Warner JR. Ribosomal protein L32 of *Saccharomyces cerevisiae* regulates both splicing and translation of its own transcript. J Biol Chem 1993; 268:19669-74.
44. Mitrovich QM, Anderson P. Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. Genes Dev 2000; 14:2173-84.
45. Umen JG, Guthrie C. Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. Genetics 1996; 143:723-39.
46. Lin R-J, Newman AJ, Cheng S-C, Abelson J. Yeast mRNA splicing in vitro. J Biol Chem 1985; 260:14780-92.
47. O'Keefe RT, Norman C, Newman AJ. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell 1996; 86:679-89.
48. Schmitt ME, Brown TA, Trumpower BL. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucl Acids Res 1990; 18:3091-2.
49. Clark TA, Sugnet CW, Ares M Jr. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 2002; 296:907-10.
50. Bousquet-Antonelli C, Presutti C, Tollervey D. Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell 2000; 102:765-75.
51. Lopez PJ, Séraphin B. YIBD: the Yeast Intron DataBase. Nucleic Acids Res 2000; 28:85-6.

6 System to Induce Rapid Protein Degradation



TIR1 from *Arabidopsis thaliana*:
Cartoon view of the 3D structure of Transport Inhibitor Response 1 protein (TIR1) from *A. thaliana* (PDB ID: 2P1P), with the cofactor inositol hexakisphosphate (IP6) and auxin. The F-box of TIR1 associates with E3 ubiquitin ligase. Auxin (in this case the analogue Indole Acetic Acid (IAA)), binds TIR1, with the co-factor IP6. Auxin stabilises the association between TIR and an Auxin Induced Degron (AID*) tag. The AID* tagged protein can then be poly-ubiquitinylated (Tan et al., 2007) and targeted for degradation.

6.1 Research Articles

This chapter is based on two peer-reviewed publications; the first covers the development and optimisation of the method and the second is a very detailed protocol.

1. Mendoza-Ochoa GI, Barrass JD, Terlouw BR, et al. (2019) A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. Yeast. 36(1):75-81.
   Figures in this publication will be referred to as P1.Figure followed by the number.
2. Barrass, J.D., Mendoza-Ochoa, G.I., Maudlin, I.E., Sani, E., Beggs, J.D., 2019. Tuning Degradation to Achieve Specific and Efficient Protein Depletion. J Vis Exp. e59874.
   This publication also includes a video of the performance of the protocol.
   Figures in this publication will be referred to as P2.Figure followed by the number.

6.2 Aim

The first publication's aim was to develop a system to rapidly and specifically degrade a target protein in yeast. Several challenges had to be overcome to produce a system that could be used reliably and flexibly in budding yeast. This resulted in the version of the Auxin Induced Degron (AID) system described in this publication.

The aim of the second publication was to facilitate the rapid and specific depletion of a protein in *S. cerevisiae* using the AID system.

6.3 Experimental System

The principal of the AID system is illustrated in Figure 1. The AID* tag is a truncated, yeast-codon-optimised AID domain derived from the plant auxin responsive protein IAA17.

TIR1, from rice (*Oriza sativa*), is an auxin receptor containing an F-box domain, by which it can associate with E3 ubiquitin ligase into a "Skp, Cullin, F-box"

complex (SCF). Both the AID* tag and TIR1 need to be introduced into yeast. The AID* tag must be fused in-frame to the coding sequence for the protein to be targeted for degradation and the TIR1 protein needs to be expressed in the same cells. Auxin, in this case the natural auxin; Indole Acetic Acid (IAA), binds to the TIR1 protein, facilitating more stable binding to the AID* tag of the target protein. This brings the target protein and the E3 ubiquitin ligase into close proximity. The target protein is then polyubiquitinated and subsequently degraded.

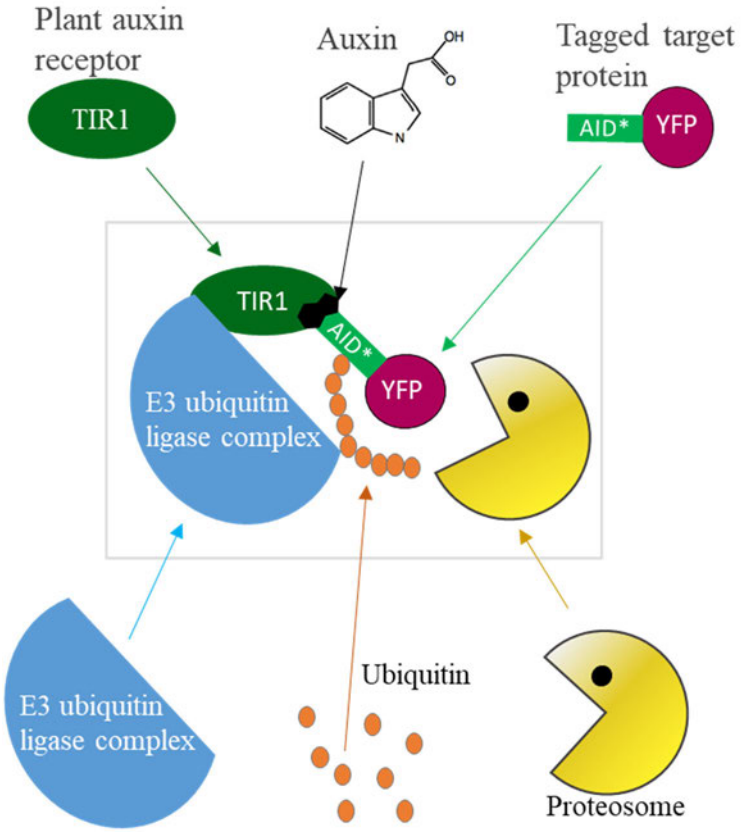


Figure 1: Basis of the AID system. The TIR1 protein and AID* tag (both in green) are derived from plants, where, in the presence of auxin, TIR1 and a protein with an AID* domain interact. TIR1 can still associate with a yeast E3 ubiquitin ligase complex which can then polyubiquitinate the tagged protein, directing it to the proteasome for degradation. YFP = Your Favourite Protein - the AID* tagged target protein.

The expression of TIR1 proved crucial to the system's correct functioning. Too little expression and depletion was poor, too much expression and the target protein was degraded without IAA addition (P1.Figure 1). The AID tag can bind to TIR1 even without auxin, albeit to a lesser extent (Tan et al., 2007). The auxin does not induce an allosteric change it merely increases the binding surface for the AID tag on TIR1.

6.3.1 Development of the Degron System

The development of the AID system is discussed in publication 1. Lower TIR1 expression is useful for partial depletion whereas strong expression is excellent for rapid and efficient depletion. If this expression is inducible (as in the β-est AID system developed in publication 1), rapid depletion is possible without lower target protein levels before depletion (see section 6.3.1.3).

6.3.1.1 GAL Promoter

The inducible gal promoter (*GAL1-10*) was originally used by the developers of the AID system to express the TIR1 protein. The strain must be shifted to galactose medium to induce expression. However, shifting from a glucose growth medium to a galactose medium severely impacts on intron splicing, so cannot be used in splicing studies (Figure 2).

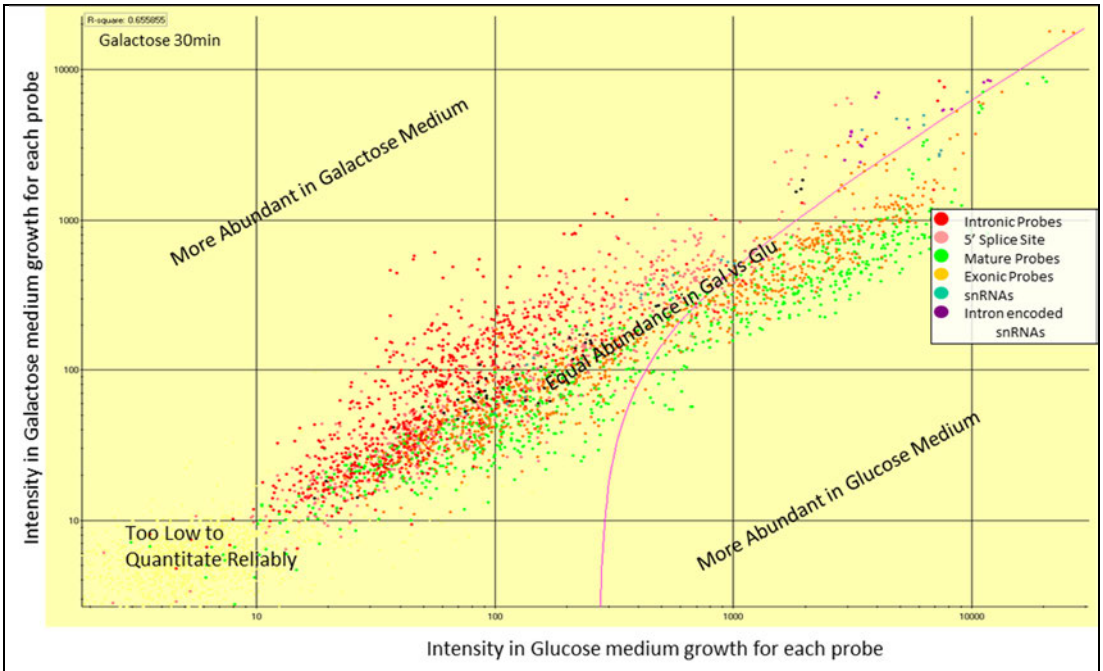


Figure 2: Effect of shift to galactose medium on pre-mRNA splicing
Microarray investigation of the effect a shift of growth medium has on splicing. The microarray design is as the introductory chapter. This is a scatterplot of the $\log_{10}$ fluorescent intensity for each probe in the microarray. The axes are the fluorescence in samples indicated. Notice the over-abundance of probes to introns (red) and 5'SS (pink) in galactose medium (YPGalA) versus glucose (YPDA). Probes to mRNA (green) are more abundant in glucose medium. This distribution indicates that splicing is affected by shift from glucose to galactose growth medium. Splicing does recover after about 12 hours in galactose medium (data not shown).

6.3.1.2 ADH1 Promoter

As a GAL promoter could not be used, two different *ADH1* promoters were trialled to express TIR1, long (701 bp) and short (409 bp) (P1.Figure 1A). Use of the long promoter results in rapid depletion of target proteins, however, the target's levels before auxin is added are lower than wild type.

The short promoter is weaker than the longer promoter and results in lower TIR1 expression and so lower depletion of even a comparatively rare protein (P1.Figure 1B & C). This is useful if a slower and less complete depletion is required. There is also no observable reduction in that protein's abundance in the absence of auxin.

6.3.1.3 Inducible Promoter, β-est AID

It is time consuming to tailor each promoter to each target protein, so an inducible promoter was used. The beta-estradiol system offered excellent induction with a small non-metabolisable chemical. The oestrogen receptor fused to the VP16 activator (also used in the ribosys reporters, chapter 2), was integrated into the genome of a yeast strain. Into this strain the promoter/TIR1 combination (shown in P1.Figure 4A), was also integrated. This system gave the required rapid depletion upon β-estradiol and IAA addition without protein depletion in their absence (P1.Figure 2).

To achieve optimum results the level of TIR1 needs to be tuned to the protein to be degraded. This is done by varying the time the TIR is expressed for before adding the auxin; the pre-incubation stage. The process for doing this is described in detail in publication 2, as correct pre-incubation is the most important factor to achieve good depletion without significant premature loss of protein.

6.3.1.4 Plasmid and NLS

The entire oestrogen and TIR1 systems were placed on a single plasmid (P2.Figure 3a). With a strain carrying this plasmid, the only modification

required is AID* tagging the target protein (P2.Figure 3b). However, the β-est AID system on a plasmid did not appear to deplete the target as well as the genomically integrated system (P1.Figure 4b & c). This is unfortunate but is the price of convenience. For a low abundance protein and with longer pre-incubation times the depletion should still be acceptable.

A nuclear localisation signal (NLS) was fused to the TIR1 used in P1.Figure 2. However, this NLS-TIR depleted a nuclear protein before the auxin was added. Presumably because the nuclear concentration of the target protein (Prp22p), was low compared to the concentration of TIR1 in the nucleus, so it was non-specifically depleted. The problem was solved by inducing TIR1 for a shorter time. Longer pre-incubation times removed the need for the NLS on TIR1 even for a nuclear protein, such as Prp22p (P1.Figure 4b). It is unclear what version of TIR (NLS or non-NLS), would be ideal for a shuttling protein, but probably both would be effective with an optimised pre-incubation.

6.4 Reinterpretation

6.4.1 Comparison with Other Methods

There are several methods that could be used to deplete a protein in yeast (Natsume and Kanemaki, 2017). One such method involves using the N-terminus rule; a temperature-sensitive peptide is fused to the N-terminus of the target protein. On shift to non-permissive temperatures the peptide's N-terminal arginine is exposed and the target protein poly-ubiquitinated and degraded. This system requires the more technically difficult N-terminal tagging and a shift to non-permissive conditions, with all the metabolic disruption entailed in a growth condition change.

Anchor-away is another frequently used system but it can only be used on nuclear proteins as this system sequesters the target protein in the cytoplasm. This sequestration is triggered by rapamycin addition, which has a major effect on yeast metabolism, triggering the starvation pathway, reducing growth and changing gene expression (Figure 3).

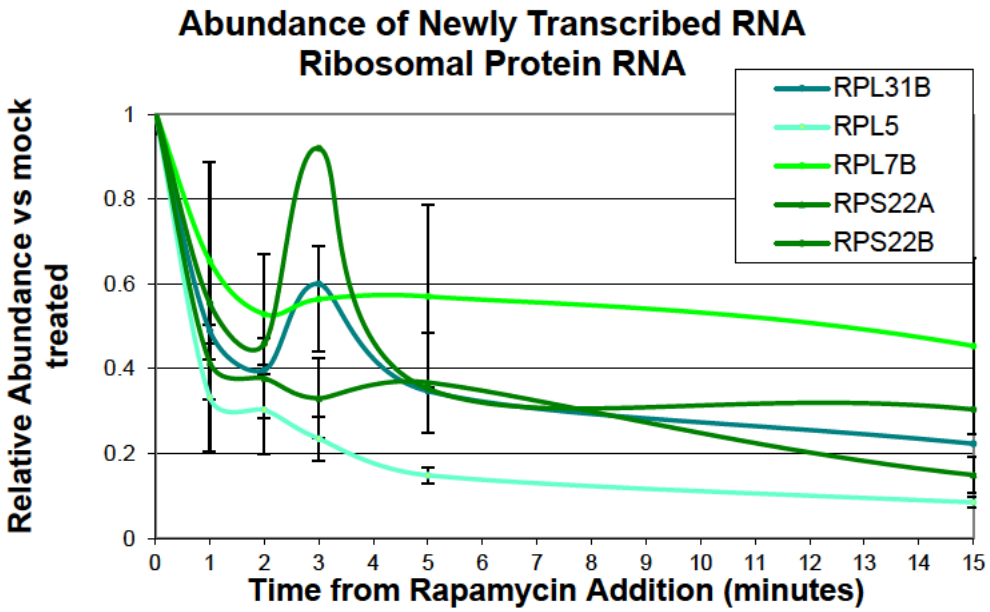


Figure 3: Effect of rapamycin on ribosomal protein gene expression
Rapamycin added to a culture, flowed by the discontinuous thio-labelling procedure as (Barrass and Beggs, 2019) (Chapter 3) Figure 2. Labelling was for 1 minute post rapamycin addition for each of the sample points. The effect on RPG expression is severe, therefore, anchor-away cannot be used in studies of splicing.

6.4.2 Limitations

While the β-est AID system is better than any other protein depletion technique published to date there are still limitations to consider.

6.4.2.1 Tagging

The target protein needs to be tagged; C-terminal tagging is comparatively easy but occasionally this disrupts the function of the protein. N-terminal tagging is possible with this system but is much more technically challenging, particularly if physiological expression levels must be maintained by using the protein's own promoter.

6.4.2.2 Auxin and β-estradiol Response

In the wild yeast grows on plants, so it is advantageous for the cell to be able to respond to plant growth. However, a wild-type strain shows no change in growth on IAA addition, whereas all AID* tagged strains tested do (Figure 4A). Similarly, β-estradiol has no effect on growth (data not shown).

Addition of auxin does induce a very mild splicing defect in wild-type yeast (Figure 4B), but the effect is slight in comparison to depletion of a splicing factor, for example Prp16p, see chapter 7 (the same scale is used in both graphs for comparison). Certainly, the splicing defect is not enough to affect growth. It is not clear what stage of splicing is affected; pre-mRNA and lariat are increased but 3'SS and mRNA levels are hardly affected. The effect could be a very mild mRNA release problem; this would increase lariat levels and also reduce the levels of available spliceosome components, retarding assembly and first step of splicing and so increase pre-mRNA levels. β-estradiol has no effect on splicing of *ACT1* pre-mRNA (Figure 4C)

6.4.2.3 Depleting 2 or 3 Proteins

Attempts were made to deplete more than one AID* tagged protein in a single strain (Figure 5). Although the western blot is poor, it is clear that the two proteins have degraded to different levels on the addition of IAA, which most likely depends on their initial abundance. This is unacceptable. Double depletion could work if the kinetics of depletion of two proteins were sufficiently similar. Depleting three proteins in a single cell was trialled with similar results, but the blot was even poorer (not shown).

6.4.3 Applications of the Degron System

The AID system, particularly β-est AID, is a very widely applicable system for depleting target proteins. If the TIR1 expression and β-estradiol receptor plasmid is used the modifications to the strain are easy to accomplish by tagging the target protein in the strain carrying the plasmid. For an example of using this system to investigate splicing see the next chapter.

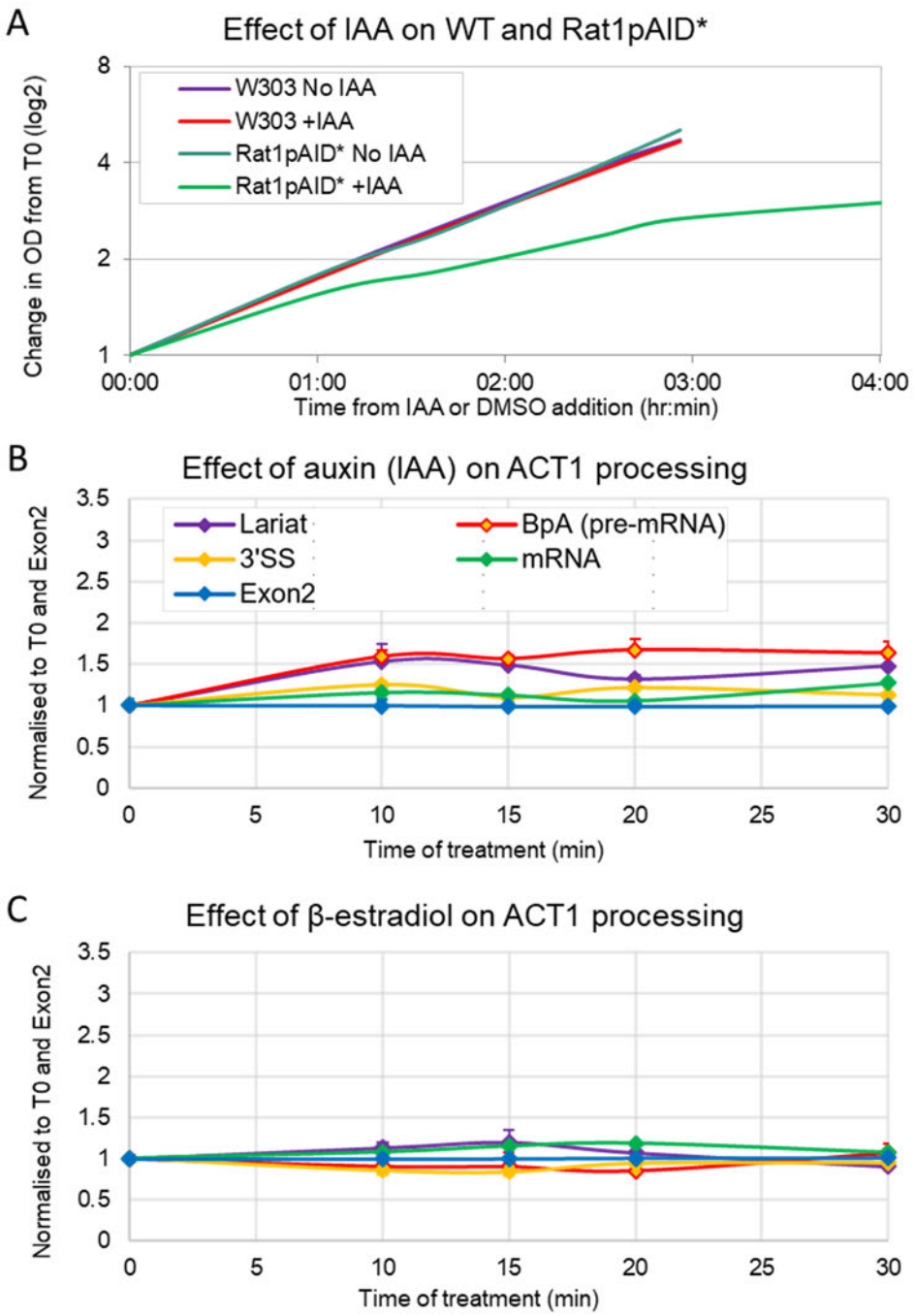


Figure 4: Effect of Auxin (IAA) and β-estradiol on yeast

A) Auxin (IAA) does not affect the growth of wild type strains but clearly has an effect on an AID* tagged strain. The tag itself has no affect on growth in this strain without IAA addition. Growth is in YPDA medium with IAA in DMSO (+IAA) or DMSO alone (No IAA). TIR expressed from the short *ADH1* promoter.

B) Splicing of *ACT1* (assayed by RT-qPCR), in the presence of auxin. Auxin does have a very slight effect on *ACT1* pre-mRNA processing, seeming to delay it.

C) Splicing of *ACT1* (assayed by RT-qPCR), in the presence of β-estradiol. β-estradiol has no effect on *ACT1* pre-mRNA splicing. Legend as B

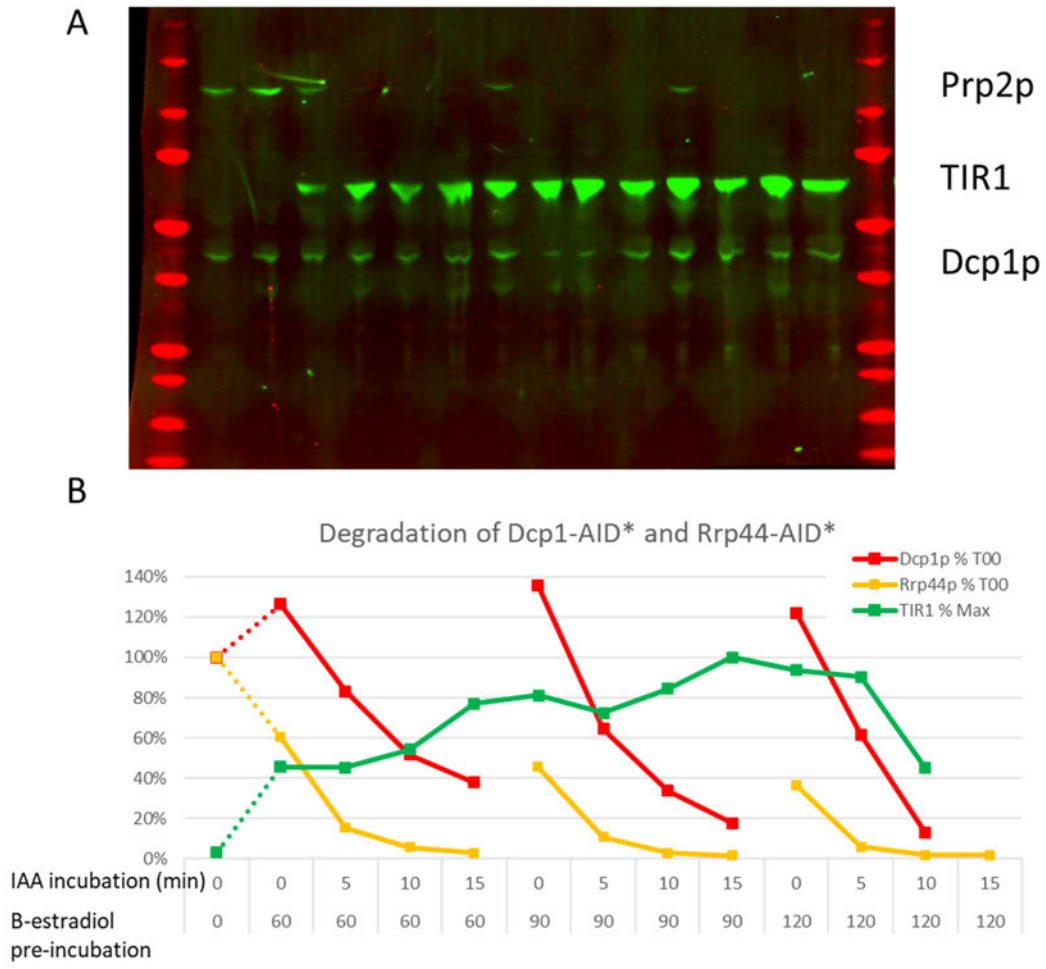


Figure 5: Depleting 2 proteins
A) Western blot (LI-COR, see both of this chapter's papers for details), showing the depletion of two proteins in one strain using the β-est AID system. These blots were at the limit of what is achievable on a single western blot with 3 antibodies. Blots with 3 proteins were attempted and, although produced results, were even poorer than the one presented here.
B) Plot of 2 different protein depletions. The initial level of Dcp1p is higher at the 0 IAA time point than the pre-incubation. This is because the first lane ran poorly. Similarly, the last sample lane is poor and the TIR1 protein could not be reliably quantitated. However, it is clear that Rrp44p (Dis3p) depleted much more than Dcp1p and started to deplete before IAA was added.

6.5 Contribution

The first publication was the result of many collaborations, of which I was a central participant. The AID project was started by Jane Reid, a PhD student

partially supervised by me, and continued by Barbara Terlouw, an honours student and then summer student intern also supervised by me. Gonzalo Mendoza, at the time a new PhD student, took up this work and did much of the later parental strain development in collaboration with me. This technique was then used by the lab on various projects. I came up with the name “β-est AID”

Publication 1; Myself and Gonzalo Mendoza are joint first authors. P1.Figures 2 and 3 are all my own work

Publication 2 was authored by me; P2.Figure 3 was Gonzalo's plasmid. Results in P2.Figure 1 are my results on my strains, the parental strain was Gonzalo's. P2.Figure 2 was the sampling system developed by me and used in many of the publications mentioned in this thesis. I perform the protocol on the video.

All additional data presented in this chapter are my own, as is any re-interpretation of the results and conclusions from the publications.

6.6 References

Most references are in the publications, a few additional ones are listed here.

Barrass, J.D., and Beggs, J.D. (2019). Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). J. Vis. Exp. e59952.

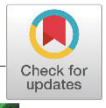
Natsume, T., and Kanemaki, M.T. (2017). Conditional Degrons for Controlling Protein Expression at the Protein Level. Annu. Rev. Genet. *51*, 83–102.

Tan, X., Calderon-Villalobos, L.I.A., Sharon, M., Zheng, C., Robinson, C.V., Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature *446*, 640–645.

6.7 Reprint 1

Received: 24 May 2018 | Revised: 2 October 2018 | Accepted: 16 October 2018

DOI: 10.1002/yea.3362



SPECIAL ISSUE ARTICLE

WILEY Yeast

A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast

Gonzalo I. Mendoza-Ochoa*,† | J. David Barrass* | Barbara R. Terlouw | Isabella E. Maudlin | Susana de Lucas | Emanuela Sani | Vahid Aslanzadeh‡ | Jane A.E. Reid§ | Jean D. Beggs

Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Correspondence
Jean D. Beggs, Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3BF, UK.
Email: j.beggs@ed.ac.uk

Present Address
†Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

‡MRC Human Genomics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

§The John Curtin School of Medical Research, The Australian National University, ACT 2600 Canberra City, Australia

Funding information
Consejo Nacional de Ciencia y Tecnología; Wellcome, Grant/Award Numbers: 092076 and 104648

Abstract

The auxin-inducible degron (AID) is a useful technique to rapidly deplete proteins of interest in nonplant eukaryotes. Depletion is achieved by addition of the plant hormone auxin to the cell culture, which allows the auxin-binding receptor, TIR1, to target the AID-tagged protein for degradation by the proteasome. Fast depletion of the target protein requires good expression of TIR1 protein, but as we show here, high levels of TIR1 may cause uncontrolled depletion of the target protein in the absence of auxin. To enable conditional expression of TIR1 to a high level when required, we regulated the expression of TIR1 using the β-estradiol expression system. This is a fast-acting gene induction system that does not cause secondary effects on yeast cell metabolism. We demonstrate that combining the AID and β-estradiol systems results in a tightly controlled and fast auxin-induced depletion of nuclear target proteins. Moreover, we show that depletion rate can be tuned by modulating the duration of β-estradiol preincubation. We conclude that TIR1 protein is a rate-limiting factor for target protein depletion in yeast, and we provide new tools that allow tightly controlled, tuneable, and efficient depletion of essential proteins whereas minimising secondary effects.

KEYWORDS

auxin, degron, estradiol, protein depletion, regulated gene expression, yeast

1 | INTRODUCTION

A common approach to study the function of an essential gene *in vivo* is to investigate the consequence of conditionally repressing its expression. Ideally, this approach should result in a fast and specific repression to minimise secondary and off-target effects. The auxin-inducible degron (AID; Nishimura, Fukagawa, Takisawa, Kakimoto, & Kanemaki, 2009) is a technique that can fulfil this goal as it allows fast depletion of the target protein by incubating the cell culture with a small molecule (auxin) that does not perturb cell metabolism. The physiological role of auxin is to regulate growth and development in plants, through a pathway that leads to the proteasome-mediated degradation of the Aux/IAA family of transcriptional regulators (reviewed in Teale, Paponov, & Palme, 2006). This pathway can be artificially transferred to a nonplant organism by expressing the TIR1 gene, which encodes a plant auxin-binding receptor that interacts with the conserved E3 ubiquitin ligase SCF complex. The target gene is fused with a transcriptional repressor Aux/IAA protein that functions as the degron tag (domain to induce degradation). When TIR1 protein is bound to auxin, it interacts with the degron, allowing the E3 ubiquitin ligase SCF complex to polyubiquitinate the target protein, thereby directing its degradation by the proteasome.

*Gonzalo I. Mendoza-Ochoa and David J. Barrass contributed equally

In the first version of the AID system for yeast, which was developed by Nishimura et al. (2009), expression of *Oryza sativa TIR1* (OsTIR1) was driven by a galactose-inducible (*GAL1–10*) promoter or the constitutive *ADH1* promoter. The galactose-inducible expression system can result in metabolic perturbations caused by the shift from glucose to galactose as carbon source (Bergkessel, Whitworth, & Guthrie, 2011; Kresnowati et al., 2006; Ronen & Botstein, 2006). Here, we tested the constitutive expression of Os*TIR1* (from here on referred to as TIR1), encoded by a codon-optimized sequence from the Kanemaki lab, under the control of a strong (P*ADH1–701*) or weak (P*ADH1–409*) promoter (Santangelo & Tornow, 1990) and show that constitutive expression of this TIR1 is problematic. High expression can result in uncontrolled target protein depletion (even in the absence of auxin). Conversely, a weak promoter can lead to slow depletion when TIR1 is expressed at low levels. This suggests that it is important to express TIR1 in a tightly controlled manner in order to achieve optimal results.

McIsaac et al. (2013) developed a β-estradiol-inducible budding yeast expression system that makes use of an artificial transcription factor, ZnEV, made by fusing a human estrogen receptor (ER), with the viral transcription activator VP16 and three or four (where n = 3 or 4) zinc finger DNA-binding domains that recognise a specific promoter, ZnEVpr. In the absence of estrogen (β-estradiol), the Hsp90 chaperone complex inhibits ZnEV by interacting with its ER domain. Binding of β-estradiol to ER releases it from the Hsp90 complex, allowing ZnEV to transcriptionally activate ZnEVpr, through direct interaction between the zinc-finger array of ZnEV and its target DNA sequence within ZnEVpr (reviewed in Pratt & Toft, 1997). McIsaac et al. (2013) report ~100-fold increase in the level of a green-fluorescent reporter protein that is produced under control of ZnEVpr only 30 min after addition of β-estradiol. More importantly, they show that the system is specific, as incubating cells with β-estradiol does not affect the global transcription profile of the yeast cells.

Here, we present a new version of the AID system for budding yeast that we call "β-est AID," in which expression of TIR1 is tightly regulated by the β-estradiol system. To test this system, we measured initial levels and depletion rate of several target proteins in a strain of *Saccharomyces cerevisiae* harbouring the β-est AID and compared them with those of strains that constitutively express TIR1. We then compared depletion rates of several nuclear target proteins following preincubation with β-estradiol for different times. The results show that depletion rate directly correlates with length of preincubation with β-estradiol and, therefore, with TIR1 levels. We demonstrate that even a highly abundant target protein can be quickly depleted by extending the preincubation with β-estradiol. Finally, we show that the β-est AID system works when the components are genomically encoded, but for ease of use, we constructed a plasmid that allows all the elements of the β-est AID system to be introduced into budding yeast through a one-step transformation.

2 | MATERIALS AND METHODS

2.1 | Plasmids

pMK200 (NBRP ID: BYP7569; Masato Kanemaki) was sourced from the Yeast Genetic Resource Centre (YGRC; Osaka University). pHyg-AID* plasmids were a gift from the Ulrich lab (Morawska & Ulrich, 2013). pBRT1 was constructed with the following DNA parts: natMX6 backbone plasmid, 409-base pair truncation of the *ADH1* promoter (P*ADH1–409*), yeast codon-optimized *TIR1* from pMK200, and *HIS3*-flanking sequences for genomic insertion. To generate plasmids pURA3-AID*-6FLAG, pURA3-AID*-9myc, and pURA3-AID*-6HA, the HPH marker in pHyg-AID* plasmids was replaced with the *Kluyveromyces lactis URA3* gene flanked by a 142 nt repeat sequence from its 5'UTR, which allows selective pop-out of the *URA3* sequence by growth in 5-fluoroorotic acid. pZTRL was constructed by Gibson Assembly with the following DNA parts: pRS415 (CEN/ARS LEU2), Z_4EVpr from pMN10 (McIsaac et al., 2013), yeast codon-optimized *TIR1* from pMK200, and *ACT1*p-Z4EV from yeast strain YMN3 (McIsaac et al., 2013). pZTRK was made by replacing the *LEU2* selection marker in pZTRL with KanMX. pMN10 was kindly provided by R. Scott McIsaac. Snapgene-generated plasmid maps (Supplemental Figure S1) and sequences of pBRT1, pURA3-AID*-6FLAG, pURA3-AID*-9myc, pURA3-AID*-6HA, pZTRK, and pZTRL can be found in the Supporting Information. These plasmids will be deposited to YGRC (http://yeast.nig.ac.jp/yeast/). The plasmid map displayed in Figure 4a was generated using Plasmid Drawing Program Plasmidomics 0.2 (Dr. Robert Winkler).

2.2 | Yeast strains and growth conditions

Yeast strains are listed in Supplemental Table S1 (These strains will be deposited to YGRC (http://yeast.nig.ac.jp/yeast/). PADH1–701-TIR1 was made by inserting *Stu*I-linearized pMK200 (containing P*ADH1–701*-Os*TIR1 URA3* expression cassette) into the *ura3–1* locus of W303, whereas PADH1–409-TIR1 was made by inserting the 4398 BP product of *Pml*1-linearized pBRT1 (which contains natMX P*ADH1–409*-Os*TIR1* expression cassette) into the *his3–11,15* locus of W303. PZ4EV-NTIR1 was created by inserting pKanMX-Z_4EVpr (from pMN10) directly upstream of the start codon of *APE2*, as described by McIsaac et al. (2013), followed by *URA3* pop-in/pop-out substitution of *APE2* protein-coding sequence for NLS-Os*TIR1*-V5. PZ4EV-TIR1 differs from PZ4EV-NTIR1 only in that it lacks the SV40 NLS in the N-terminus of *OsTIR*-V5. *PRP22*, *PRP2*, *DCP1*, *YHC1*, and *RRP44* were AID*-tagged by transforming PADH1–701-TIR1, PADH1–409-TIR1, PZ4EV-NTIR1, PZ4EV-TIR1, or pZTRL-bearing W303, with pHyg-AID*- or pURA3-AID*- cassettes, following a PCR-based method (Longtine et al., 1998). YMN3 was kindly provided by R. Scott McIsaac. Yeast were grown at 30°C on Yeast Peptone Dextrose supplemented with adenine (YPDA) or yeast minimal media (YMM) supplemented with Kaiser drop outs (Formedium). When required, β-estradiol (Sigma-Aldrich) was added at a final concentration of 10 uM and indole-3-acetic acid (auxin; Acros Organics) at a final concentration of 750 uM. Samples of yeast cultures were fixed in methanol, chilled in dry ice at a ratio of 3:2 (culture:methanol by volume).

2.3 | Western blots

Protein extracts were prepared following a NaOH/TCA precipitation method (Volland, Urban-Grimal, Géraud, & Haguenauer-Tsapis, 1994). Protein concentrations were measured by the Bradford method

and equal amounts of protein were loaded for each sample into SDS-PAGE. Proteins were transferred to Immobilon-FL PVDF (Millipore, Cat. No. IPFL00010) and probed with rat anti-FLAG (Agilent, Cat. No. 200474), mouse anti-HA (Roche, Cat. No. 11583816001), mouse anti-PGK1 (Abcam, Cat. No. Ab113687) and/or mouse anti-V5 (Invitrogen, Cat. No. MA5–15253); and LI-COR secondary antibodies (Cat. No. 925–32219, 925–68020 and/or 925–32210). Rabbit anti-OsTIR1 antibody was provided by Masato Kanemaki. Blots were developed with the LI-COR Odyssey system and images analysed using the Odyssey Image Studio software to quantify the band signals. Normalisation was by amount of protein loaded. Pgk1 bands were also quantified for comparison.

2.4 | Reverse transcription-quantitative polymerase chain reaction

RT-qPCR was performed as described in (Alexander et al., 2010). Primers for RT-qPCR were TTGCTGCAAGATTCCCAAACG (TIR1 167F), CCCC AATCAGGTGGAACCAA (TIR1 254R), TAAGCTGGCATGTGCTGCATTC (ALG9_F), and TTTGCATGATTCGGTTGATTGG (ALG9_R). Transcript copy numbers per cell were estimated by comparing the relative abundance of TIR1 amplicon to that of internal control *ALG9* (1.28 copies/cell; Miura et al., 2008; Teste, Duquenne, François, & Parrou, 2009).

3 | RESULTS

3.1 | High-level expression of TIR1 causes auxin-independent depletion of target protein

First, we compared strains PADH1–701-TIR1 and PADH1–409-TIR1 that constitutively express TIR1 to high or low levels, respectively (Figure 1a,b). To this end, we C-terminally tagged splicing factor Prp22 with a fusion between a truncated auxin-dependent degron, named AID* (Morawska & Ulrich, 2013) and six tandem repeats of the FLAG epitope, in strains PADH1–701-TIR1 and PADH1–409-TIR1. To measure the rate of target protein depletion, we added auxin (indole-3-acetic acid) to the cultures (time 0), and samples were taken after 5, 15, and 30 min and snap frozen.

Quantification of proteins in the samples (Figure 1b,c) shows that at time 0 (no auxin), the levels of Prp22 were 47% lower in PADH1–701-TIR1 than in PADH1–409-TIR1. As PADH1–701-TIR1 constitutively expresses TIR1 to a higher level, this may indicate that too much TIR1 can cause uncontrolled depletion of the target protein. Also, the Prp22 depletion rate was higher in PADH1–701-TIR1 than in PADH1–409-TIR1. By 30 min after auxin addition, Prp22 levels had dropped to below 6% of the initial values in PADH1–701-TIR1, compared with 35% remaining in PADH1–409-TIR1, suggesting that high levels of TIR1 promote faster depletion.

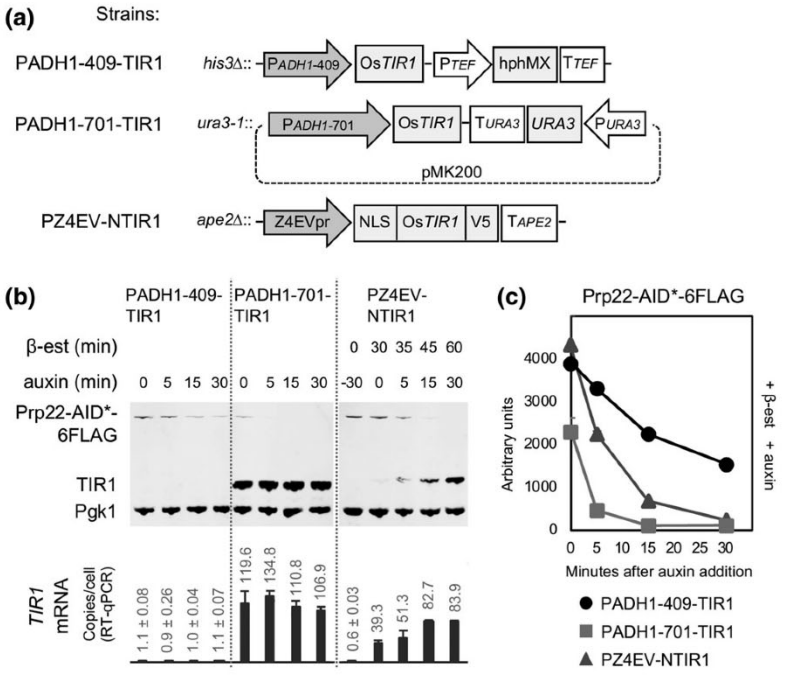


FIGURE 1 High-level expression of TIR1 causes auxin-independent depletion of target protein. (a) genomically integrated *TIR1*-expression cassettes of strains, PADH1–409-TIR1, PADH1–701-TIR1, and PZ4EV-NTIR1 used in this study. Z4EVpr promoter is induced by β-estradiol (β-est) whereas P*ADH1* promoters are constantly active. (b) Upper: Western blot of Prp22-AID*-6FLAG, TIR1, and Pgk1 (as visual loading control; equal amounts of total protein were loaded in each lane) at different times (min; minutes) of β-estradiol and/or auxin incubation; lower: estimated copies/cell of *TIR1* transcripts from the corresponding culture samples. (c) quantification of western blot shown in panel B

We, therefore, created an inducible-TIR1 strain by placing the *OsTIR1* gene under control of the Z4EV promoter (Z4EVpr). This was done by replacing the nonessential *APE2* gene in the Z4EV-expressing strain YMN3 (McIsaac et al., 2013) with a Z4EVpr-Os*TIR1*-V5 cassette (Figure 1a). To promote localisation of TIR1 protein to the nucleus, where our protein targets are located, we fused an SV40 nuclear localisation signal (NLS) to the *TIR1*-coding sequence. The resulting strain, PZ4EV-NTIR1, produces TIR1-V5 protein rapidly after addition of β-estradiol to the culture medium (Figure 1b), and auxin was added after preinduction of TIR1 for 30 min. Under these conditions, Prp22 was depleted rapidly after auxin addition, without detectable auxin-independent depletion (Figure 1b,c).

Next, to test the hypothesis that levels of TIR1 inversely correlate with levels of the target protein in the absence of auxin, a culture of PZ4EV-NTIR1 with AID*-6FLAG-tagged *PRP22* was incubated with β-estradiol but without auxin, and the levels of Prp22 were measured over time. Consistent with our previous observation, at 50 min of β-estradiol incubation, the level of Prp22 had dropped significantly and reached 35% of the initial value at 2 hr of incubation, by which time the TIR1-V5 protein was well induced (Figure 2). Auxin-independent depletion of Yhc1 and Rrp44 is shown in Figure S2 and was less pronounced with the more abundant Rrp44. We conclude that high levels of TIR1 can cause auxin-independent depletion of the target protein in budding yeast.

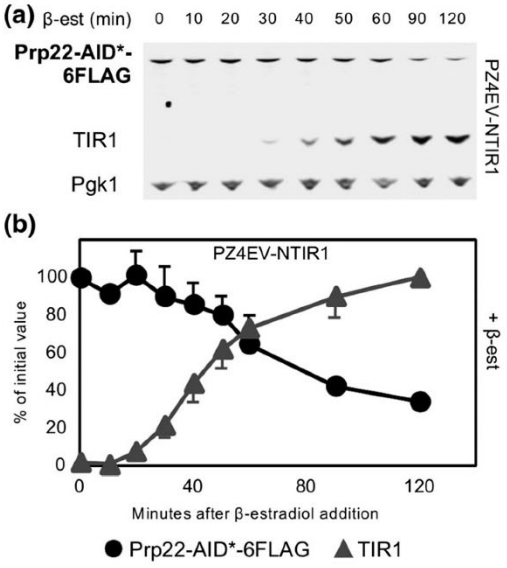


FIGURE 2 Auxin-independent depletion can be controlled by inducible expression of TIR1. (a) Western blot of AID-target Prp22-AID*-6FLAG and β-estradiol-inducible TIR1 (strain PZ4EV-NTIR1), after addition of β-estradiol but without auxin. (b) Quantification of western blot shown in panel A. Error bars represent standard deviation (sd) of two biological replicates. For each experiment, only one representative blot is shown. Error bars represent standard deviation (sd) of two biological replicates. For each experiment, only one representative blot is shown

3.2 | Depletion rate can be tuned by modulating the duration of β-estradiol preincubation

Next, we investigated how the rate of auxin-induced depletion is influenced by the length of β-estradiol preincubation. On the basis of the previous results, we anticipated that not only longer preincubation times would lead to faster depletion but also more auxin-independent depletion and that there may be an optimal preincubation time, which is likely to be target specific. To test this, we used as targets for depletion, Prp22 (232 copies/cell), Prp2, another essential splicing factor with similar abundance (211 copies/cell), and the more highly expressed decapping enzyme, Dcp1 (4,189 copies/cell; Kulak, Pichler, Paron, Nagaraj, & Mann, 2014). We performed a time-course depletion analysis in which cultures of these AID*-tagged strains were preincubated with β-estradiol for different times (20, 30, 40, or 60 min) prior to auxin addition. Samples were then taken for protein analysis at 5-min intervals. As the levels of TIR1 increase with time of β-estradiol incubation, this allowed us to measure the relationship between TIR1 abundance and auxin-dependant depletion rate of different target proteins.

The protein quantification analysis (Figure 3), shows that different target proteins were depleted at different rates. A 20-min preincubation with β-estradiol was sufficient to reduce Prp22 and Prp2 to low levels (≤ 20%) within 15 min of auxin addition, but longer preincubations with β-estradiol resulted in auxin-independent degradation. In contrast, the more abundant Dcp1 required 60-min of β-estradiol preincubation to achieve a similarly rapid and efficient depletion because more Dcp1 protein has to be degraded to achieve efficient depletion in terms of percentage of the starting amount, and this evidently requires more TIR1 protein. Notably, with all three target proteins, there was a direct correlation between the duration of β-estradiol preincubation and the depletion rate so that the duration of β-estradiol treatment should be optimised for each target protein (see also Figure S2).

3.3 | Plasmid-encoded β-est AID and the effect of TIR1 protein localization on efficiency of targeted depletion

To facilitate insertion of the β-est AID components into budding yeast, we constructed a centromeric plasmid, pZTRL, which contains P*ACT1*-Z4EV and Z4EVp-Os*TIR1* (without NLS) expression cassettes (Figure 4 a). We then tested pZTRL and, at the same time, investigated the effect of fusing an NLS to TIR1 on the efficiency of depleting a nuclear protein. To this end, we measured both TIR1 and Prp22 target protein levels in the pZTRL-bearing strain, and in two strains that contain genomically integrated *TIR1*, with (PZ4EV-NTIR1) or without (PZ4EV-TIR1) NLS on the N-terminal of TIR1 (Figure 4b,c).

We observed that the β-estradiol-induced level of TIR1 protein was about threefold higher in the genomic *TIR1* (PZ4EV-TIR1) strain compared with genomic NLS-*TIR1* (PZ4EV-NTIR1) or plasmid-encoded *TIR1* strains, indicating that *TIR1* is better expressed when it is genomically-integrated and lacks an NLS. Interestingly, Prp22, which is a nuclear localized protein, was quickly depleted irrespective of the

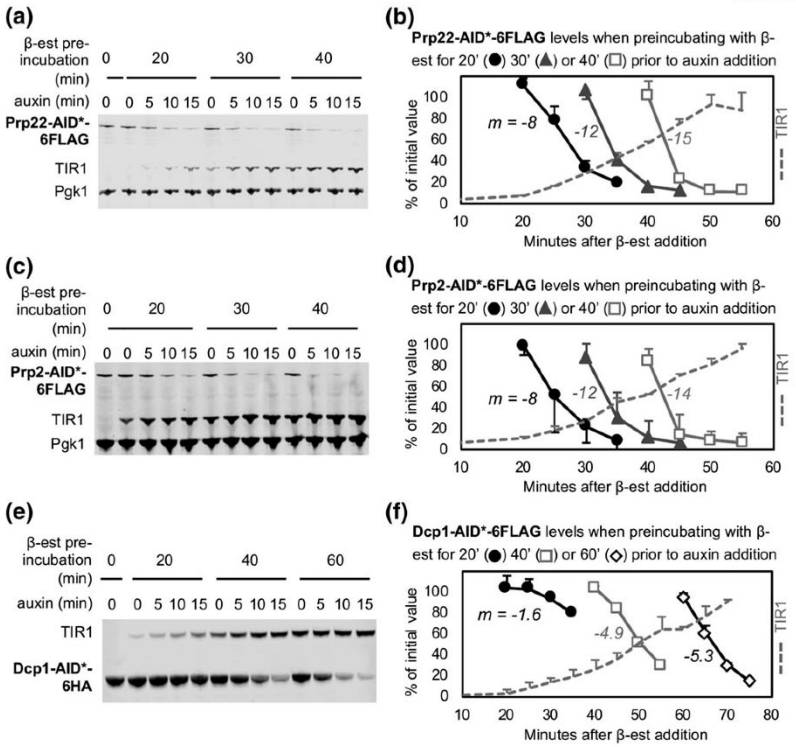


FIGURE 3 Depletion rate can be tuned by modulating the duration of β-estradiol pre-incubation. Western blot of AID-targets (a–b) Prp22-AID*-6FLAG, (c–d) Prp2-AID*-6FLAG, and (e–f) Dcp1-AID*-6HA, from cultures preincubated with β-estradiol (β-est) for 20, 30, 40, or 60 min prior to auxin addition. Equal amounts of total protein were loaded in each lane, and Pgk1 is included as a visual loading control, except for panel E, where Pgk1 and Dcp1 comigrate. Quantifications of protein bands in panels A, C, and E are shown in panels B, D, and F. As a measure of depletion rate, the slope (m) was calculated for the linear section (from 100 to 30% of initial values) of each curve. Error bars represent sd of two biological replicates. For each experiment, only one representative blot is shown

presence or absence of the NLS in the N-terminus of TIR1 protein. This suggests that, for depletion of nuclear proteins, an SV40 NLS need not be added to TIR1. Notably, targeted depletion was slower in the plasmid-encoded TIR1 strain compared with the other two strains, reaching 15% compared with 2% of Prp22 initial values 30 min after auxin addition, probably due to a lower expression of TIR1 in this strain.

4 | DISCUSSION

Our observation that in budding yeast high levels of TIR1 protein cause auxin-independent depletion agrees with reports that over-expression of TIR1 in *Arabidopsis thaliana* leads to an auxin-response phenotype (Gray et al., 1999) and depletion of TIR1 substrates (Aux/IAA proteins; Dos Santos Maraschin, Memelink, & Offringa, 2009), even without exogenous auxin. More recently, Natsume, Kiyomitsu, Saga, and Kanemaki (2016) constructed a tetracycline-inducible TIR1 for human cells and showed that incubation with the tetracycline analogue doxycycline slows growth of a DHC1-AID tagged cell culture, implying that in human cells, over-expression of TIR1 may cause leaky depletion of the target (Natsume et al., 2016). In our work, we were repeatedly unsuccessful with 10 out of 20 essential genes that we tried to AID-tag in PADH1–701-TIR1 (high-level TIR1 expression), whereas eight of these were successfully tagged in either PADH1–409-TIR1 (low-TIR1 expression) or in pZTRL. Indeed, all 22 essential genes for which AID tagging was attempted in PADH1–409-TIR1 (low-TIR1 expression), and all nine attempts to AID-tag essential genes in pZTRL were successful (data not shown), with success defined as failure to grow in the presence of auxin. This, together with our measurements of Prp22 and TIR1 levels in these strains, strongly suggests that high-level expression of TIR1 causes auxin-independent degradation that may result in target proteins falling below the level required for viability. The auxin independent activity of SCF-TIR1 E3 ubiquitin ligase complex that we and others have observed could be caused by low-affinity interaction of TIR1 with its target in the absence of auxin (Dharmasiri, Dharmasiri, & Estelle, 2005; Kepinski & Leyser, 2005; Tan et al., 2007); by traces of auxin in the media, as previously speculated (Natsume & Kanemaki, 2017); or by low levels of endogenous auxin in plants and yeast (Rao, Hunter, Kashpur, & Normanly, 2010).

Notably, the plasmid-encoded, β-est AID (pZTRL) depleted the target protein more slowly than genomically integrated TIR1 (Figure 4), likely because the plasmid-encoded TIR1 protein was

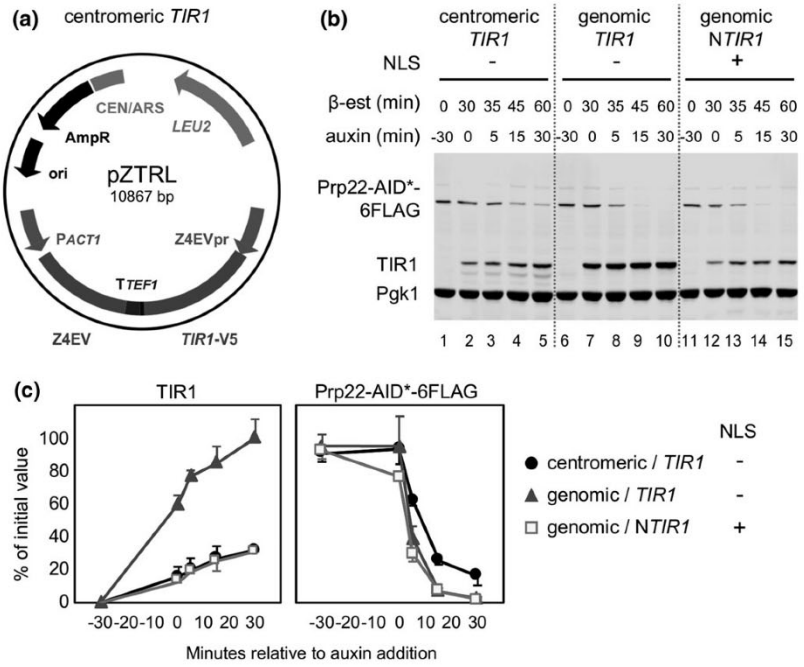


FIGURE 4 Plasmid-encoded β-est AID and the effect of TIR1 protein localization on targeted depletion efficiency. (a) Plasmid map of pZTRL. (b) Western blot of TIR1 and AID-target Prp22-AID*-6FLAG in three different strains preincubated with β-estradiol for 30 min (min) prior to auxin addition. Equal amounts of total protein were loaded in each lane, and Pgk1 is included as a visual loading control. In the first strain (lanes 1–5), both Z4EVpr-Os*TIR1*-V5 and P*ACT1*-Z4EV expression cassettes are located in pZTRL centromeric plasmid; whereas in strains PZ4EV-TIR1 (lanes 6–10) and PZ4EV-NTIR1 (lanes 11–15), they are genomically encoded. In PZ4EV-NTIR1, an SV40 nuclear localization signal was included in the N-terminal of *TIR1*. (c) Quantification of western blot signal shown in panel B. Error bars represent sd of two biological replicates. Only one representative blot is shown

produced at a lower level. This could be due to transcriptional interference between P*ACT1*-Z4EV and Z4EVpr-Os*TIR1* that are convergently transcribed in pZTRL, for example, if their shared transcriptional terminator T*TEF1* does not act bidirectionally—even though it was proposed that most transcriptional terminators in *S. cerevisiae* function bidirectionally (Uwimana, Collin, Jeronimo, Haibe-Kains, & Robert, 2017). Thus, the plasmid-based β-est AID may require longer β-estradiol preincubation times and may not be ideal for protein targets that are more abundant. In addition to pZTRL, we created pZTRK by replacing *LEU2* in pZTRL with KanMX. pZTRK can be propagated in rich media, which is useful given that high concentrations of auxin can impair growth in minimal media (T.S. Turowski, S. Bresson and D. Tollervey, personal communication).

We targeted only nuclear proteins for depletion, which is why we tested the effect of an NLS sequence at the start of the TIR1 protein. However, as previously reported for budding yeast (Tanaka, Miyazawa-Onami, Lida, & Araki, 2015), the presence of the NLS had little effect on depletion rate of a nuclear protein (Prp22). We speculate that TIR1 protein when N-terminally fused to an NLS may have reduced stability, but that this may result in similar amounts of nuclear localised TIR1 protein in the presence or absence of an NLS, such that the target protein depletion rates are similar. Nishimura et al. (2009) showed that the AID system allows the rapid and efficient depletion of proteins present either in the nucleus or the cytoplasm. Therefore, we anticipate that our non-NLS TIR1 would also support efficient depletion of cytoplasmic targets although we have not tested this.

In summary, we present evidence that the level of expression of TIR1 has a profound influence on how rapidly and completely the AID system depletes its target protein. In order to enhance this system for the molecular biology community, we developed a suite of strains and plasmids in *S. cerevisiae*. The strain PADH1–409-TIR1 allows a slower, more linear depletion of the target protein (Figure 1c), which is useful where full depletion is undesirable, and may facilitate kinetic studies of the effects of protein depletion. The β-estradiol-induced TIR1 strains can have their preincubation time optimised to produce an extremely rapid and effective degradation of the target protein whereas minimising auxin-independent degradation. Although this tunable β-estradiol expression system has been developed specifically for use in budding yeast, the data and principles we present are likely to apply to other organisms and to be useful for the wider scientific community interested in making the most out of the powerful AID technique.

ACKNOWLEDGEMENTS

We are grateful to Scott McIsaac for providing pMN10 and strain YMN3, Helle Ulrich for pHyg-AID* plasmids, and Masato Kanemaki

for the gift of rabbit anti-OsTIR1 antibody and for helpful discussion. This work was supported by a scholarship to GIMO from the Consejo Nacional de Ciencia y Tecnología, Mexico (CONACYT) and the University of Edinburgh School of Biological Sciences, a Darwin Trust of Edinburgh studentship to VA, Wellcome 4-year PhD studentships [093853] to JEAR, and [105256] to IEM, and by Wellcome funding [104648] to JD Beggs. Work in the Wellcome Centre for Cell Biology is supported by Wellcome core funding [092076].

ORCID

Jane A.E. Reid http://orcid.org/0000-0002-5196-8121
Jean D. Beggs http://orcid.org/0000-0002-8179-6519

REFERENCES

Alexander, R. D., Barrass, J. D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M. C., … Beggs, J. D. (2010). RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. *RNA*, *16*, 2570–2580. https://doi.org/10.1261/rna.2162610

Bergkessel, M., Whitworth, G. B., & Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. *RNA*, *17*, 1461–1478. https://doi.org/10.1261/rna.2754011

Dharmasiri, N., Dharmasiri, S., & Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature*, *435*, 441–445. https://doi.org/10.1038/nature03543

Dos Santos Maraschin, F., Memelink, J., & Offringa, R. (2009). Auxin-induced, SCF^{TIR1}-mediated poly-ubiquitination marks AUX/IAA proteins for degradation. *The Plant Journal*, *59*, 100–109. https://doi.org/10.1111/j.1365-313X.2009.03854.x

Gray, W. M., Pozo, J. C., Walker, L., Hobbie, L., Risseeuw, E., Banks, T., … Estelle, M. (1999). Identification of an SCF ubiquitin–ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes & Development*, *13*, 1678–1691.

Kepinski, S., & Leyser, O. (2005). The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature*, *435*, 446–451. https://doi.org/10.1038/nature03542

Kresnowati, M. T. A. P., Van Winden, W. A., Almering, M. J. H., Ten Pierick, A., Ras, C., Knijnenburg, T. A., … Daran, J. M. (2006). When transcriptome meets metabolome: Fast cellular responses of yeast to sudden relief of glucose limitation. *Molecular Systems Biology*, *2*, 49.

Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods*, *11*, 319–324. https://doi.org/10.1038/nmeth.2834

Longtine, M., McKenzie, A., Demartini, D., Shah, N., Wach, A., Brachat, A., … Pringle, J. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, *14*, 953–961.

McIsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. A., Botstein, D., & Noyes, M. B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research*, *41*, e57. https://doi.org/10.1093/nar/gks1313

Miura, F., Kawaguchi, N., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y., & Ito, T. (2008). Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics*, *9*, 574.

Morawska, M., & Ulrich, H. D. (2013). An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast*, *30*, 341–351. https://doi.org/10.1002/yea.2967

Natsume, T., & Kanemaki, M. T. (2017). Conditional degrons for controlling protein expression at the protein level. *Annual Review of Genetics*, *51*, 83–102. https://doi.org/10.1146/annurev-genet-120116-024656

Natsume, T., Kiyomitsu, T., Saga, Y., & Kanemaki, M. T. (2016). Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Reports*, *15*, 210–218. https://doi.org/10.1016/j.celrep.2016.03.001

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, *6*, 917–922. https://doi.org/10.1038/nmeth.1401

Pratt, W. B., & Toft, D. O. (1997). Steroid Receptor Interactions with heat shock protein and immunophilin chaperones. *Endocrine Reviews*, *18*, 306–360. https://doi.org/10.1210/edrv.18.3.0303

Rao, R. P., Hunter, A., Kashpur, O., & Normanly, J. (2010). Aberrant synthesis of indole-3-acetic acid in *Saccharomyces cerevisiae* triggers morphogenic transition, a virulence trait of pathogenic fungi. *Genetics*, *185*, 211–220. https://doi.org/10.1534/genetics.109.112854

Ronen, M., & Botstein, D. (2006). Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 389–394. https://doi.org/10.1073/pnas.0509978103

Santangelo, G. M., & Tornow, J. (1990). Efficient transcription of the glycolytic gene ADH1 and three translational component genes requires the GCR1 product, which can act through TUF/GRF/RAP binding sites. *Molecular and Cellular Biology*, *10*, 859–862. https://doi.org/10.1128/MCB.10.2.859

Tan, X., Calderon-Villalobos, L. I. A., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., & Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, *446*, 640–645. https://doi.org/10.1038/nature05731

Tanaka, S., Miyazawa-Onami, M., Lida, T., & Araki, H. (2015). iAID: An improved auxin-inducible degron system for the construction of a "tight" conditional mutant in the budding yeast *Saccharomyces cerevisiae*. *Yeast*, *32*, 567–581. https://doi.org/10.1002/yea.3080

Teale, W. D., Paponov, I., & Palme, K. (2006). Auxin in action: Signalling, transport and the control of plant growth and development. *Nature Reviews. Molecular Cell Biology*, *7*, 847–859. https://doi.org/10.1038/nrm2020

Teste, M. A., Duquenne, M., François, J. M., & Parrou, J. L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Molecular Biology*, *10*, 99. https://doi.org/10.1186/1471-2199-10-99

Uwimana, N., Collin, P., Jeronimo, C., Haibe-Kains, B., & Robert, F. (2017). Bidirectional terminators in *Saccharomyces cerevisiae* prevent cryptic transcription from invading neighboring genes. *Nucleic Acids Research*, *45*, 6417–6426. https://doi.org/10.1093/nar/gkx242

Volland, C., Urban-Grimal, D., Géraud, G., & Haguenauer-Tsapis, R. (1994). Endocytosis and degradation of the yeast uracil permease under adverse conditions. *The Journal of Biological Chemistry*, *269*, 9833–9841.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Mendoza-Ochoa GI, Barrass JD, Terlouw BR, et al. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast*. 2018;1–7. https://doi.org/10.1002/yea.3362

6.8 Reprint 2

jove Journal of Visualized Experiments www.jove.com

Video Article

Tuning Degradation to Achieve Specific and Efficient Protein Depletion

J. David Barrass[1], Gonzalo I. Mendoza-Ochoa[1,2], Isabella E. Maudlin[1,3], Emanuela Sani[1], Jean D. Beggs[1]

[1]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh

[2]Department of Plant Sciences, University of Cambridge

[3]Sir William Dunn School of Pathology, University of Oxford

Correspondence to: J. David Barrass at david.barrass@ed.ac.uk

URL: https://www.jove.com/video/59874
DOI: doi:10.3791/59874

Keywords: Immunology and Infection, Issue 149, AID, auxin, β-estradiol, degron, protein depletion, regulated gene expression, *Saccharomyces cerevisiae*

Date Published: 7/20/2019

Citation: Barrass, J.D., Mendoza-Ochoa, G.I., Maudlin, I.E., Sani, E., Beggs, J.D. Tuning Degradation to Achieve Specific and Efficient Protein Depletion. *J. Vis. Exp.* (149), e59874, doi:10.3791/59874 (2019).

Abstract

The plant auxin binding receptor, TIR1, recognizes proteins containing a specific auxin-inducible degron (AID) motif in the presence of auxin, targeting them for degradation. This system is exploited in many non-plant eukaryotes, such that a target protein, tagged with the AID motif, is degraded upon auxin addition. The level of TIR1 expression is critical; excessive expression leads to degradation of the AID-tagged protein even in the absence of auxin, whereas low expression leads to slow depletion. A β-estradiol-inducible AID system was created, with expression of TIR1 under the control of a β-estradiol inducible promoter. The level of TIR1 is tunable by changing the time of incubation with β-estradiol before auxin addition. This protocol describes how to rapidly deplete a target protein using the AID system. The appropriate β-estradiol incubation time depends on the abundance of the target protein. Therefore, efficient depletion depends on optimal timing that also minimizes auxin-independent depletion.

Video Link

The video component of this article can be found at https://www.jove.com/video/59874/

Introduction

Conditional mutations, such as temperature-sensitive mutants, are a powerful tool for the study of essential proteins, allowing cell growth under the permissive condition but causing loss of function under non-permissive conditions. However, cell metabolism can be seriously perturbed by the change in growth conditions required to induce the defect and may also create off-target effects. Several methods have been developed, in which the protein of interest is conditionally sequestered[1] or its expression is controlled[2,3] by addition of a small molecule. This protocol uses auxin and the auxin-inducible degron (AID) system to efficiently deplete a target protein.

The AID system has its origin in plants, where an auxin (in this protocol indole-3-acetic acid (IAA) is used), stimulates interaction of the Aux/IAA protein with TIR1, a member of the SCF U3 ubiquitin ligase complex[4]. SCF complex interaction causes polyubiquitination of Aux/IAA family proteins, which results in their degradation by the proteasome[5,6].

This system was previously adapted for use in the yeast *Saccharomyces cerevisiae*[7,8] by expressing the TIR1 protein from *Oriza sativa* (osTIR) in yeast cells, where it is able to interact with the endogenous yeast SCF complex. The protein of interest was tagged with a motif from the Aux/IAA protein IAA17 to target it for degradation. Functional truncations of IAA17 were developed later, such as AID*[8,9,10], containing the 43 amino acid auxin-sensitive motif from *Arabidopsis thaliana IAA17*, along with an epitope tag to enable detection.

The system initially adapted for use in budding yeast[7,8] expressed the osTIR1 protein from a yeast GAL promoter. Expression requires shifting to growth medium with galactose as the sole carbon source, which, unfortunately, results in a diauxic shift with wide-ranging changes to cell metabolism[11]. On the other hand, it has been reported that constitutive expression of TIR1 can lead to degradation of the target protein in the absence of auxin/IAA[12] if the expression level is high, whereas low TIR1 expression causes inefficient depletion. An improved AID system named β-est AID was developed in which the osTIR is under the control of an inducible promoter that is tunable to suit the target protein, with minimal effect on cell metabolism. To achieve this, an artificial transcription factor (ATF) was constructed in which the *VP16* viral transcription activator is fused to an oestrogen receptor and a four Zn fingers DNA binding domain (DBD). When β-estradiol (an oestrogen) is present, the ATF can enter the nucleus and induce osTIR transcription by binding to its promoter (Z4EVpr)[13,12].

osTIR expression is usually detectable about 20 min after addition of β-estradiol[12]. However, the optimal duration of osTIR expression to achieve efficient depletion of the tagged protein with auxin, while avoiding depletion before auxin addition, needs to be empirically determined for each target protein. An approximate time for this pre-incubation can be estimated from abundance values in the Saccharomyces Genome Database (SGD https://www.yeastgenome.org/). As can be seen in **Figure 1**, the abundant protein, Dcp1 (2880 to 4189 molecules/cell), requires 40 min of pre-incubation with β-estradiol, with no auxin-independent depletion observed. The much less abundant protein, Prp2 (172 to 211 molecules/cell), is strongly depleted after only 20 min of pre-incubation. It is advisable to test two additional pre-incubation times, 10 to 20 min before or

jove Journal of Visualized Experiments www.jove.com

after this initial estimated time (20 min is the minimum time that is recommended). The optimum pre-incubation time is the time at which target protein has not depleted before adding auxin and once auxin is added the depletion is acceptable or protein levels approach the minimum possible. So, from **Figure 1b**, for Prp22 with 30 min of pre-incubation, the levels have not declined much 10 min after auxin addition. Comparing this with 40 min of pre-incubation and 15 min with IAA, where there is little additional depletion, there is no benefit in incubating with auxin longer than 10 min or pre-incubating for longer than 30 min, particularly as there is evidence of non-auxin dependent depletion at 40 min. For Dcp1 with 40 min of pre-incubation (the last point at which the protein level is approximately 100% before auxin addition), 15 to 20 min of depletion with auxin is acceptable. It is recommended to keep the depletion time as short as possible to reduce secondary effects on cell metabolism[14].

This article demonstrates how to use the β-est AID system by optimizing the timing of β-estradiol incubation for osTIR expression to achieve rapid target protein depletion upon IAA addition without depletion before adding auxin.

Protocol

NOTE: See **Figure 2** for a graphical summary.

1. Strain Preparation

1. **Using a *ura3*-strain, introduce the β-est AID system (i.e., genes encoding the β-estradiol responsive transcription factor (ATF) and the osTIR) and AID* tag the target protein (see Figure 3 and Table 1 for a summary of the procedure).**
    1. Transform[15] either pZTRK (G418 resistance marker) or pZTRL (*LEU2* marker) plasmid (available from the Yeast Genetic Resource Centre) into the *ura3*- yeast strain or use the plasmid as a template to produce the PCR product for genomic integration.
    2. PCR amplify the ATF (marked Z4EVATF on the plasmid map) and osTIR using a high fidelity polymerase from either of the plasmids pZTRK or pZTRL. Use primers with 50 to 60 base 3' extensions with homology to the genomic region, to direct integration by homologous recombination[16]. For genomic integration of the two components either separately or together, see **Table 1** for primers and conditions.
    NOTE: The strain pZ4EV-NTR1 has the components already integrated in the genome (available from the Yeast Genetic Resource Centre, Japan).
    3. Ensure that the target protein is AID* tagged using the Longtine procedure[17] (see **Figure 3b** and **Table 1**).
    4. Perform a growth analysis on the strain without β-estradiol and IAA present to determine if the AID* tag affects growth and to predict growth rate for use at step 1.5.

2. General Procedure for Depletion

1. Calculate how much culture is required for all samples to be collected; for example, 10 mL of culture at OD_{600} of 0.8 is sufficient for protein, RNA and DNA extraction for a single sample, so for 6 samples, at least 60 mL of culture is needed.
2. From an overnight culture, set up sufficient new culture at OD_{600} 0.1 to 0.2 and leave to grow at 30 °C. A rich medium such as YPDA is recommended, although other growth conditions can be used:

Yeast Extract	10 g
Peptone	20 g
Glucose	20 g
Adenine sulphate	40 mg
H_2O to	1 L

NOTE: Autoclave or filter sterilize; filter sterilization is preferred as peptide/sugar complexes produced by autoclaving precipitate in the methanol used in sample collection.

3. **Prepare to receive the samples.**
    1. Put 30 to 50% of the intended sample volume of methanol into a tube. For example, if a 10 mL sample is to be taken, put 5 mL of methanol into a 15 mL falcon tube and close the tube tightly. Once closed, label the tube and put on dry ice or at -80 °C to chill.
    CAUTION: Dispense the methanol in a fume hood.
    2. Label 1.5 mL tubes for long term storage of the samples and place in ice to cool.
    3. Cool enough H_2O (at least 1 mL per sample) on ice.
4. Anticipate the culture's growth. The target OD for collecting the samples is approximately 0.7 to 0.8, but the pre-incubation step (the incubation with β-estradiol to induce the osTIR), needs to be started earlier so that the culture will reach approximately the right OD by the time the samples are collected.
NOTE: It is advisable to perform a growth curve in the conditions to be used in the experiment so that this starting OD can be estimated.
5. Once the target OD for the start of the pre-incubation has been reached, take a sample (usually 10 mL), into the pre-prepared tube containing cold methanol. Invert briefly to mix and place back in dry ice.
NOTE: The sample can be moved to water ice after about 5 min, if convenient to do so.
6. Immediately add the β-estradiol, 1 µL/mL of culture (final concentration of 10 µM); have the β-estradiol pre-measured in a pipette ready for use in order to reduce the time taken between collecting the sample and adding the β-estradiol. Rapidly mix by swirling vigorously.
7. Continue to grow the culture as before (step 2.2), incubate (this is the "pre-incubation" step) with β-estradiol for the optimal time (for determination of the optimal pre-incubation time see **Figure 1**).
8. Prepare to add IAA (auxin). Take up the volume of IAA needed for step 2.10 (i.e., 0.5 µL of IAA per mL of culture). This makes step 2.20 faster.

jove Journal of Visualized Experiments www.jove.com

9. Collect a sample as step 2.5.
10. Immediately add IAA 0.5 µL/mL of culture to a final concentration of 750 µM as prepared in step 2.8. Rapidly mix by swirling vigorously.
11. Collect samples, as step 2.5, according to your experimental design. Either a single sample, at a time when it is expected that the protein will be reliably depleted, or multiple samples in a time course of depletion. For example, 5 min intervals are convenient for timing and provide a range of protein levels. The optimization strategy, as shown in **Figure 1**, will give an indication of suitable times.
12. **Process the samples.**
    1. Place the samples on ice, if not done already. Ensure that none of the samples has frozen; if they have, gently warm in the hand, inverting constantly so the temperature does not rise locally.
       NOTE: This is best done in the hand as the sample's temperature can be assessed, it should always feel cold. Place on ice. This is not a pause point - once all the samples are fluid, proceed to the next step.
    2. Once all samples have been collected and are no longer frozen, spin at 3,500 x *g* for 2 min (at 4 °C if possible).
    3. Pour off the methanol/medium mix and place back on ice; do not worry if not all the liquid has been removed.
    4. Resuspend the cell pellet in 1 mL of ice cold H_2O (from step 2.3.3) and transfer to a labelled 1.5 mL tube (prepared in step 2.3.2) on ice.
    5. Spin briefly (e.g., 10 s total time) at >15,000 x *g* to re-pellet the cells, place back on ice and remove the liquid.
    6. Remove the H_2O by aspiration. The cell pellets can be stored at -20 °C, or -80 °C for long term storage.
13. Check the level to which the protein has been depleted by Western blot analysis[18].
    NOTE: Sufficient protein[19] and/or nucleic acid can be extracted from a single cell pellet for most purposes, although rare RNA species might require more sample volume.

Representative Results

Representative examples of depletion are displayed in **Figure 1**. The three experiments presented in this figure were optimization experiments for depletion of the proteins Prp2, Prp22 and Dcp1. The low abundance, spliceosomal Prp2 and Prp22 proteins both depleted to less than 20% after 40 min pre-incubation with β-estradiol followed by 15 min with auxin. Longer pre-incubation times lead to faster depletion but also show undesirable protein depletion before auxin addition. In comparison, the more abundant Dcp1 was only depleted to approximately 30% with the same treatment, but 60 min of pre-incubation resulted in depletion to 13% with the same auxin treatment, at the cost of depletion before the auxin is added. It is possible that 50 min of pre-incubation with β-estradiol and 15 min with auxin would have achieved similar results at a shorter time point and so would have been more optimal.

jove Journal of Visualized Experiments www.jove.com

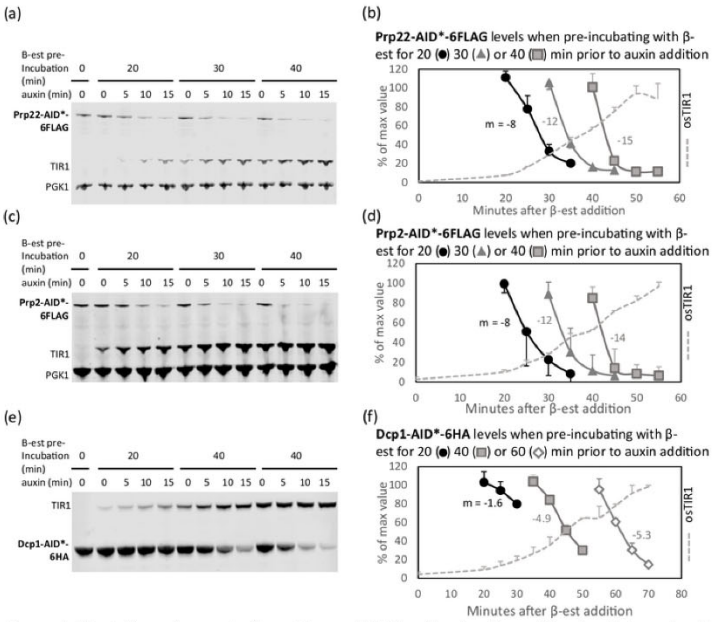


Figure 1: Depletion rate can be tuned by modulating the duration of β-estradiol pre-incubation. Western blot[18] of target proteins: (**a** and **b**) Prp22-AID*-6FLAG, (**c** and **d**) Prp2-AID*-6FLAG, and (**e** and **f**) Dcp1-AID*-6HA, from cultures pre-incubated with β-estradiol (β-est) for 20, 30, 40, or 60 min prior to auxin addition[12]. Equal amounts of total protein were loaded in each lane. Pgk1 is detected as a visual loading control, except for panel **e**, where Pgk1 and Dcp1 co-migrate. Quantification of protein bands in panels **a**, **c** and **e** are shown in panels **b**, **d**, and **f**, respectively. As a measure of depletion rate, the slope (m) was calculated for the linear section (from 100% to 30% of initial values) of each curve. The optimal pre-incubation time is the time at which the protein levels are still close to the un-induced levels (100%) and the subsequent rate of depletion is fast. For Dcp1 (**f**), 60 min of pre-incubation is too long, as the protein has begun to degrade in the absence of auxin, whereas 20 min is too short, as the protein does not appreciably deplete in this time course. After 40 min pre-incubation, 15 min with auxin can be used as the protein is approximately 70% depleted and, although 20 min would result in further depletion, it could also result in secondary effects. Error bars represent standard deviation of two biological replicates. For each experiment, one representative blot is shown. This figure is derived from previous publication[9]. Please click here to view a larger version of this figure.

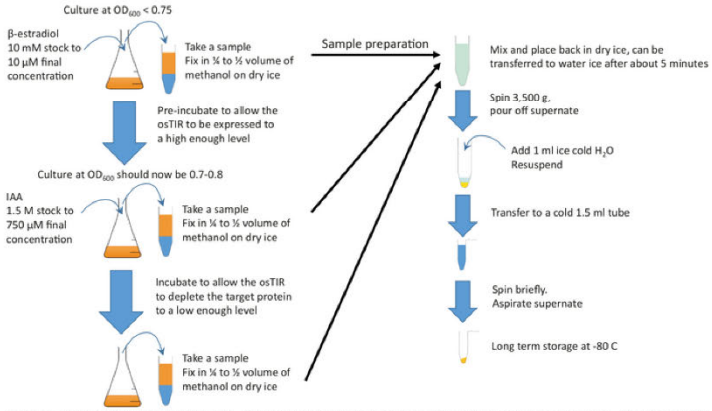


Figure 2: Graphical summary. Add β-estradiol to sufficient culture growing in rich medium and at the required temperature in order to start pre-incubation. Continue growth for the required pre-incubation time before adding IAA (auxin) to start depletion. The pre-incubation and depletion times depend on the protein to be depleted, but pre-incubation is often in the range of 20-60 min and the depletion time is typically in the order of 10 to 20 min. 10 mL samples should be taken at the start and end of pre-incubation and during the depletion. These samples are rapidly fixed in cold methanol before pelleting and storage. Please click here to view a larger version of this figure.

jove Journal of Visualized Experiments www.jove.com

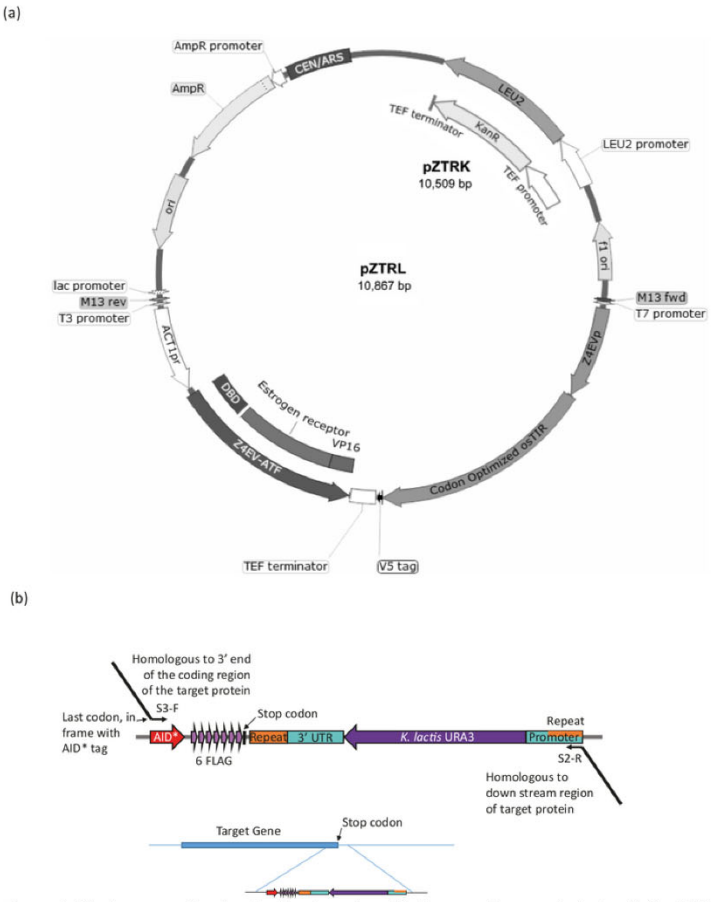


Figure 3: Strain generation for the B-est system. (**a**) To generate a yeast strain with the AID* system, either the pZTRL (*LEU2*) or the pZTRK (kanamycin (G418) resistance), plasmid should be introduced into the strain or, alternatively, the ATF and osTIR can be inserted into the genome by homologous recombination of a fragment generated by PCR from 3'end primers (see **Figure 3b** and **Table 1**). (**b**) C-terminal tagging of the target protein is achieved by PCR amplification of the appropriate region of the plasmid pURA3-AID*-6FLAG (pURA3_AID*-6HA differs only in the tag and can be treated in exactly the same way), using Longtine primers S3-F and S2-R with 3' extensions homologous to the 3' end of the target protein. The forward primer extension should include the last amino acid codon in frame with the start of the AID* tag and must not include the stop codon. The reverse primer extension should be to a region downstream of the coding region. Once inserted into the genome, cells that have lost the *URA3* marker (by homologous recombination between the identical regions found at both ends of the marker) can be selected by growth with 5-FOA, that counter-selects *URA3* cells. Please click here to view a larger version of this figure.

a. Primer Sequences

Target	Location	Primer	Name	Sequence	Tm (°C)
pZTRL	516	F	pZTRL_F	<-region of homology->GCGACAGCATCACCGACTTCG	61.23
	7897	R	pZTR_R	CGCCGCCTCTACCTTGCAGA<-region of homology (RC)->	61.30
pZTRK	9154	F	pZTRK_F	<-region of homology->ACGTTGAGCCATTAGTATCAATTTGCTTACC	59.40
	5897	R	pZTR_R	CGCCGCCTCTACCTTGCAGA<-region of homology (RC)->	61.30
pURA3-AID*-6FLAG or pURA3_AID*-6HA		F	S3-F	<-region of homology->CGTACGCTGCAGGTCGAC	59.21
		R	S2-R	ATCGATGAATTCGAGCTCG<-region of homology (RC)->	52.76

pZRTL/K is to amplify the β-est AID system	
pURA3-AID*-6FLAG/6HA to amplify the AID* and epitope tag to tag the target protein (Lontine procedure)	
<-region of homology->	Region homologous to the flanking regions where the system is to be inserted. The longer this region is the more likely the modification is to be successful; 50 - 100 bases is recommended.
<-region of homology (RC)->	Region homologous to the flanking regions where the system is to be inserted, remember to use the reverse complement. As above, the longer this region is the better.
Tm (°C)	Tm using the %GC method with 50 mM NaCl

b. PCR Mix

Component	Volume (µL)
Template	<10
NEB Phusion HF Buffer (5x)*	100
Forward Primer 100 µM	2.5
Reverse Primer 100 µM	2.5
dNTPs 10 mM each	10
H_2O	to 500

* The NEB Phusion GC Buffer (5x) can also be used but is not preferred
Make this mix, split into 10 tubes of 50 µl mix each and perform the PCR as Table 1 c.
Check the PCR has worked by running on an agarose gel
Combine all successful reactions into one tube and ethanol precipitate
Transform the yeast with all the material produced by the PCR

c. PCR Conditions

Step		Temp (°C)	Time
Initial Denaturation		98	30 s
25-35 Cycles	Denature	98	10 s
	Anneal	45–60	20 s
	Extension	72	30 s/kb
Final Extension		72	10 min
Hold		8	

Anneal at 45 °C for the Lontine primer set (S3-F and S2-R) and 60 °C for the pZTRL/K primers
Extend for 3 minutes for the Lontine PCR and 3 minutes for pZTRL/K

Table 1: Primer sequences, PCR mix and PCR conditions.

jove Journal of Visualized Experiments www.jove.com

Discussion

A well optimized protocol can produce rapid and efficient depletion of the target protein. Determining the approximate pre-incubation time with β-estradiol is important, as this increases reproducibility of the depletion, but small variations in pre-incubation time can be tolerated. On the other hand, care must be taken with timing after auxin addition, as the protein level declines very rapidly.

An advantage of this approach is that tuned depletion can be achieved by varying combinations of pre-incubation time with β-estradiol and IAA incubation time. For example, if desired, the target protein can be more slowly depleted by reducing the pre-incubation time.

The β-est AID system offers certain advantages over systems where OsTIR is constitutively expressed. For example, if the target protein is essential for viability, regulated expression of osTIR can avoid premature depletion of the target protein. Moreover, expression of osTIR can be tuned to suit the abundance of the target protein and its susceptibility to degradation, and the depletion can be either fast or slow. The two small molecule effectors, β-estradiol and auxin, do not perturb the yeast metabolism under the conditions used here, unlike rapamycin, used in the anchor-away system[1].

It should be noted that tagging some proteins disrupts their function, which is a problem with any targeted depletion system. In this case, an N-terminal tag may work when a C-terminal tag does not. Also, not all proteins will be depleted efficiently; for example, the AID-tag on the target protein may be inaccessible to the osTIR protein. Therefore, after AID-tagging, each target protein should be tested for any effect of the tag on growth, and to determine whether depletion is effective, before the timings of β-estradiol pre-incubation and auxin treatment are optimized.

This AID* system is very simple and is compatible with any subsequent experimental procedure that does not involve further growth, such as protein, DNA or RNA analysis or microscopy. In addition, the system works well when combined with thiolabelling to purify nascent RNA[20].

This system provides a rapid, specific, and reproducible means of depleting a protein without otherwise affecting the metabolism of the yeast cell.

Disclosures

The authors have nothing to disclose.

Acknowledgments

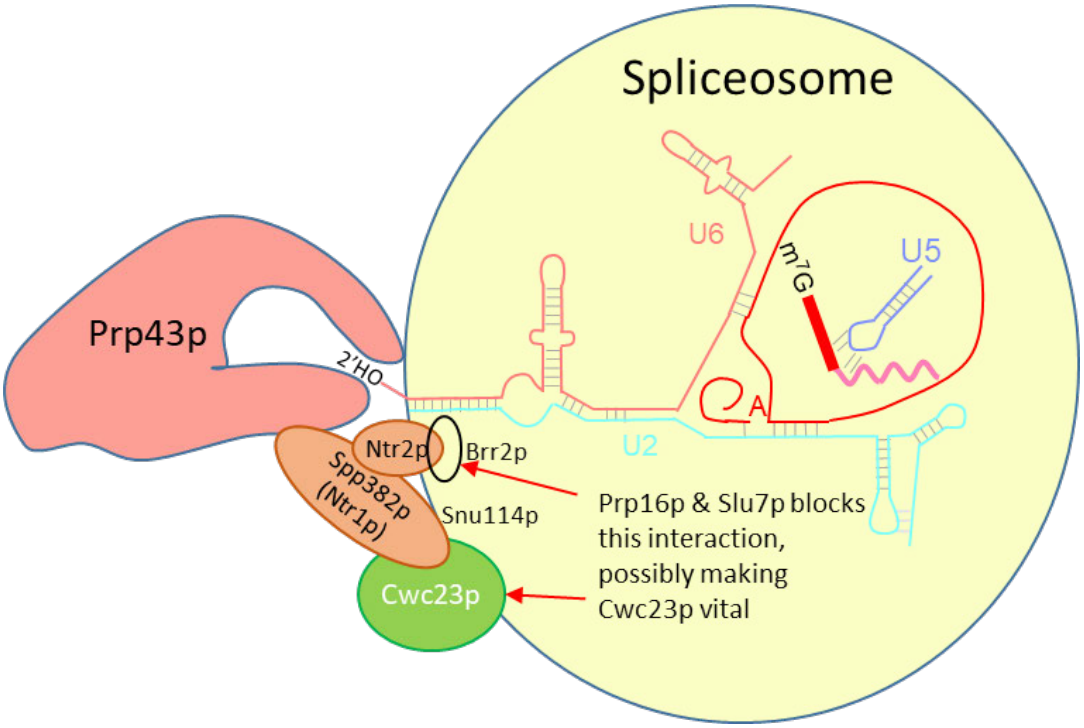
Thanks to Jane Reid for initiating this programme, Barbara Terlouw for development, Vahid Aslanzadeh for the "ura looper" constructs and Susana de Lucas for many helpful discussions. This work was supported by a scholarship to GIMO from the Consejo Nacional de Ciencia y Tecnología, Mexico (CONACYT) and the University of Edinburgh School of Biological Sciences, a Wellcome PhD studentship to IEM [105256] and by Wellcome funding [104648] to JD Beggs. Work in the Wellcome Centre for Cell Biology is supported by Wellcome core funding [092076].

References

1. Haruki, H., Nishikawa, J., & Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Molecular Cell.* **31**, 925–932 (2008).
2. Bellí, G., Garí, E., Piedrafita, L., Aldea, M., & Herrero, E. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Research.* **26**, 942–947 (1998).
3. Alexander, R. D. *et al.* RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in Saccharomyces cerevisiae. *RNA.* **16**, 2570–2580 (2010).
4. Deshaies, R. J., & Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annual Review of Biochemistry.* **78**, 399–434 (2009).
5. Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature.* **446**, 640–645 (2007).
6. Teale, W. D., Paponov, I. A., & Palme, K. Auxin in action: signalling, transport and the control of plant growth and development. *Nature Reviews Molecular Cell Biology.* **7**, 847–859 (2006).
7. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods.* **6**, 917–922 (2009).
8. Morawska, M., & Ulrich, H. D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast.* **30**, 341–351 (2013).
9. Kubota, T., Nishimura, K., Kanemaki, M. T., & Donaldson, A. D. The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication. *Molecular Cell.* **50**, 273–280 (2013).
10. Brosh, R. *et al.* A dual molecular analogue tuner for dissecting protein function in mammalian cells. *Nature Communications.* **7**, 11742 (2016).
11. DeRisi, J. L., Iyer, V. R., & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science.* **278**, 680–686 (1997).
12. Mendoza-Ochoa, G. I. *et al.* A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast.* (2018).
13. McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57 (2013).
14. Prusty, R., Grisafi, P., & Fink, G. R. The plant hormone indoleacetic acid induces invasive growth in Saccharomyces cerevisiae. *PNAS.* **101**, 4153–4157 (2004).
15. Geitz, D., St Jean, A., Woods, R. A., & Schiest, R. H. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Research.* **20**, 1425 (1992).
16. Widlund, P. O., & Davis, T. N. A high-efficiency method to replace essential genes with mutant alleles in yeast. *Yeast.* **22**, 769–774 (2005).
17. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast.* **14**, 953–961 (1998).

18. Eaton, S. L. *et al.* A Guide to Modern Quantitative Fluorescent Western Blotting with Troubleshooting Strategies. *Journal of Visualized Experiments* e52099 (2014).
19. Volland, C., Urban-Grimal, D., Géraud, G., & Haguenauer-Tsapis, R. Endocytosis and degradation of the yeast uracil permease under adverse conditions. *Journal of Biological Chemistry.* **269**, 9833–9841 (1994).
20. Barrass, J. D. *et al.* Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biology.* **16**, 282 (2015).

7 Using Protein Degradation to Disrupt Spliceosome Recycling



Prp43p unpicks the spliceosome:
The helicase Prp43p is required to disassemble the spliceosome. It binds the 2' OH at the 3' end of U6 and is presumed to pull on the RNA, unravelling the spliceosome (Toroney et al., 2019). The other components of the NTR complex; Spp382p (formerly known as Ntr1p), and Ntr2p assist. Cwc23p is needed to dissociate stalled spliceosomes, possibly because splicing factors are blocking the association Ntr2p normally has with Brr2p in the post splicing spliceosome (Su et al., 2018). According to the timing model of quality control (see thesis introduction) the final timer for stalled spliceosome disassembly would be the time taken for Cwc23p, Ntr2p, Spp382p and Prp43p to assemble on the spliceosome, followed by ATP hydrolysis before the spliceosome can be disassembled.

7.1 Research Article

This chapter is based on the peer-reviewed publication: Mendoza-Ochoa, G.I., Barrass, J.D., Maudlin, I.E., Beggs, J.D., 2019. Blocking late stages of splicing quickly limits pre-spliceosome assembly in vivo. RNA Biol 16, 1775–1784.

Figures in this publication will be referred to as P.Figure followed by the number.

This publication describes an investigation of splicing using the methods in the publications discussed in all the previous chapters.

7.2 Aim

The AID* system (see chapter 6), was used to rapidly and specifically deplete two late acting spliceosomal helicases. This depletion was performed using both rapid and slower AID* systems in order to distinguish primary effects on splicing from secondary effects. The primary effects are caused directly by depleting the helicases and secondary effects are mediated by spliceosome depletion arising from defects in spliceosomal component recycling.

7.3 Experimental System

As can be seen from the splicing diagram (Figure 5, thesis introduction) the spliceosome components are recycled for new rounds of splicing. If a step late in splicing or spliceosome disassembly is disrupted spliceosomes become trapped on transcripts, unable to complete splicing; this is the primary effect. Consequently, the levels of free spliceosome components fall, so early stages of splicing are then affected, a secondary effect. The experiments in this publication combine thiolabelling (chapters 3 & 4) with AID (particularly β-est AID) (chapter 6) to create time-based data similar to the Ribosys project (chapter 2) and produce insights into splicing similar to chapter 5.

The two factors depleted using the AID system are Prp16p and Prp22p. Both are helicases and, as discussed in the introduction, have a role in error checking the step before in which they act; Prp16p is a second step factor and checks step 1 of splicing (see introduction). Prp22p acts later in the splicing

reaction, the release of the spliced product, and checks step 2. For comparison, two other factors, Prp4p and Prp45p, acting in spliceosome assembly, were also depleted.

7.3.1 Depletion of Splicing Factors Affects Splicing

Protein depletion and its effects are shown in P.Figure 1. P.Figure 1a shows the level of protein depletion compared to un-depleted and panel d shows the effect of depletion on splicing of *ACT1*, measured by the RT-qPCR system expounded in the introduction. As can be seen, depletion of any of these four splicing factors leads to accumulation of pre-mRNA. Pre-mRNA can only accumulate if the efficiency of first step of splicing is reduced or is blocked entirely. A naive explanation is that these are all first step splicing factors. However, upon Prp16p and Prp22p depletion levels of lariats also rise, whereas they decline when the factors, Prp4p and Prp45p, are depleted.

An explanation is that Prp16p and Prp22p are indeed second step splicing factors but, having failed to complete step 2 or release the mRNA, the spliceosomes are trapped. Pre-mRNA builds up because a late stage problem prevents the recycling of spliceosome components, therefore the next generation of pre-mRNA transcripts cannot be spliced.

7.3.2 Spliceosome Assembly

If the above proposed model is correct effects on spliceosome assembly should be seen by ChIP (P.Figure 2). ChIP with U1 and U2 factors (Prp40p and Lea1p respectively), is remarkably consistent given the different stages the depleted factors act on the spliceosome. Where there is no depletion; U1 (black line), peaks over the 3'SS and 5' end of exon2. When the splicing factors are depleted (dashed line) more U1 builds up and persists later into the second exon, except Prp4p depletion.

Prp4p is a member of the tri-snRNP, so U1 binding should be unaffected. However, when Prp4p is depleted U1 levels marginally decline over the 3'SS and start of exon 2. A possible explanation is that the U1 is sequestered on transcripts where the U1 has not been replaced by U6 because Prp4p is not

present. Normally the U1 is released to bind nascent transcripts again, but without Prp4p it becomes trapped in failed spliceosomes. If this is the case, this is a recycling defect of the U1 snRNP.

U2 in un-depleted strains peaks over the body of the gene (red line). On depletion of any of the proteins tested the U2 snRNP signal is reduced and the peak of detection shifts further downstream in exon 2 (dashed red line). All the depleted proteins join the spliceosome after U2 binding and so cannot influence U2 directly. The ChIP results indicate that in the absence of these factors the spliceosome fails to assemble properly past the commitment complex (E1). The poorer U2 detection in the depleted samples indicates that U2 is sequestered away from the spliceosome.

These results are reinforced by the RNA immuno-precipitation experiments (RIP), P.Figure 3. The U2 component, Lea1p, is precipitated and the U1, U2, U4, U5 and U6 snRNAs assayed by RT-qPCR. In the un-depleted sample (P.Figure 3a) Lea1 is mostly associated with the later spliceosome, i.e. after U4 has dissociated from the spliceosome. By comparison in Prp4p depletion (P.Figure 3b) the distribution shows U2 snRNP associated only with U1, in the commitment complex. On Prp45p depletion more U1 is detected implying Lea1 is in spliceosomes at an early stage of assembly. Prp16p and Prp22p depletion traps U2 in more complete spliceosomes. U2 snRNP accumulates in stages of spliceosome assembly prior to the point of action of the proteins depleted, indicating the spliceosome is stalled by the protein's absence.

7.3.3 Kinetic ChIP

All the previous depletion experiments were done after 30 minutes of depletion. For P.Figure 4b ChIP of U2 was performed during a time course of Prp22p depletion. On both the *ACT1* and *ECM33* genes U2 signals declined during Prp22p depletion. U2 levels on *ACT1* took several minutes to decline, as would be expected if U2 was initially assembled correctly on the transcript but was not available for later transcripts due to sequestration in failed spliceosomes. The kinetics of *ECM33* U2 loss are similar but there is no lag in

decline of U2. A possible explanation is that the nuclear location of *ECM33* has a smaller buffer of free U2 than *ACT1*'s neighbourhood.

7.3.4 Thiolabelling During Depletion

Using the discontinuous thiolation procedure, outlined in chapter 3's publication ((Barrass and Beggs, 2019) Figure 2), the efficiency of splicing was monitored as Prp16p was depleted using β-est AID. Thio-labelling was for one minute at regular intervals throughout the depletion time course. The results are shown in P.Figure 4c and more clearly in (Barrass and Beggs, 2019) Figure 3c and reproduced here in Figure 1b. This shows an initial build-up of *ACT1* lariats from 4 to 6 minutes post depletion but after 10 minutes they fall below steady state levels. 10 minutes is also the time at which pre-mRNA, (Branch Point, (BpA) PCR product), significantly increases. This would indicate that by 4 minutes (about 35% of normal Prp16p levels), second step is compromised but step 1 is not yet significantly affected. From 6 minutes there are fewer and fewer lariats being made by the first step as spliceosome components are trapped in stalled spliceosomes. By 10 minutes the level of one or more of the spliceosome components required for spliceosome assembly is limiting, first step of splicing fails; no lariats are generated and the pre-mRNA builds up. Spliceosomes must be being disassembled as the lariats decline after 6 minutes, but the spliceosome components are not available for spicing, either the components are not completely liberated from the stalled complex or are degraded.

The percent of Prp16p present is also indicated by the black line in Figure 1b. However, splicing does not co-vary with the protein level, see Figure 1c as it does with time. This suggests that the splicing defect seen in Figure 1 reflects stocks being depleted through time rather than the protein having different roles, revealing themselves at different concentrations.

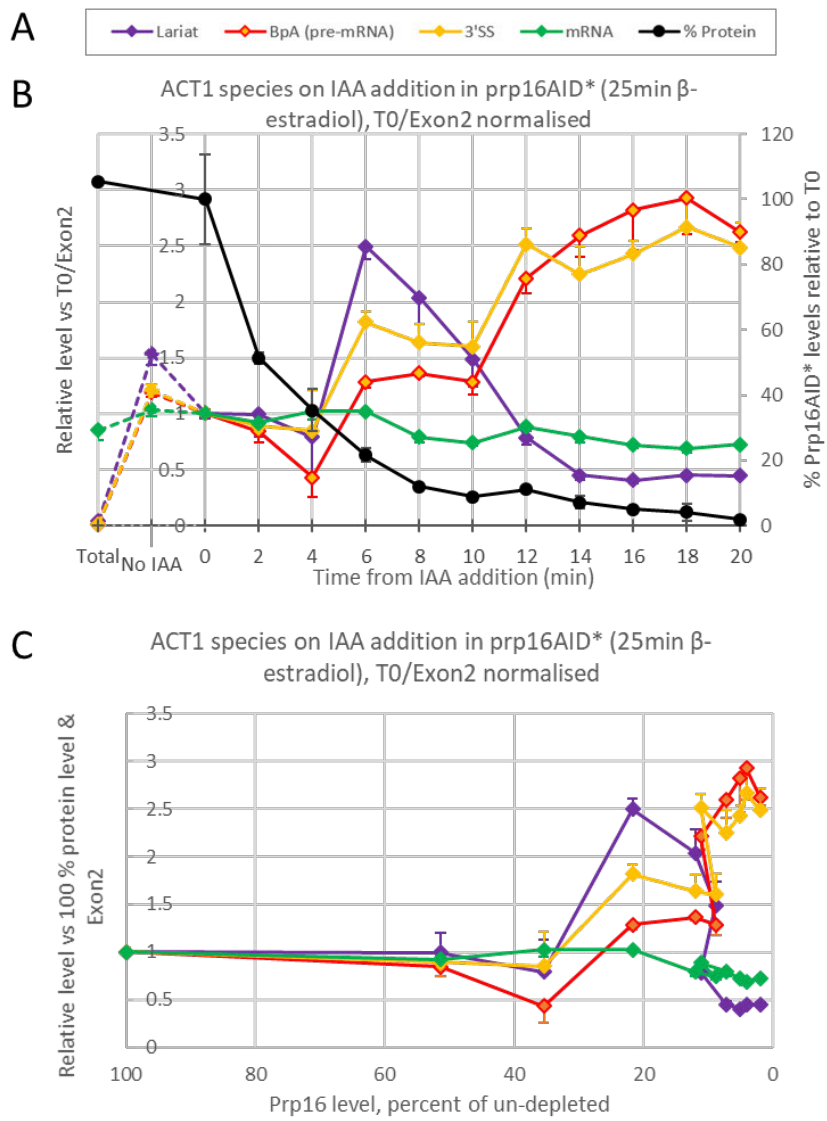


Figure 1: *ACT1* RNA species vary in abundance as Prp16p is depleted
The *ACT1*; pre-mRNA, lariat and mRNA levels normalized against the level of Exon2 and steady state levels of these RNAs and how they change as Prp16AID* is depleted.
A Key to Plots
B Splicing of ACT1 pre-mRNA transcripts upon depletion of the Prp16 protein as plotted against the length of time of depletion. The Prp16 protein levels are also displayed in the graph plotted against the second Y-axis as percentage of levels prior to depletion. Lariats accumulate initially as the second step of splicing is affected. At later time points splicing fails completely, lariats are no longer produced and pre-mRNA levels rise. Reproduced from Barrass and Beggs (2019)
C As B but the RNA abundance is plotted against the protein levels. There is no effect on splicing until Prp16p levels are <50% of normal, an indication of the spare capacity of Prp16p in the cell during log-phase growth at the *ACT1* locus. Once there is only about 15% of the protein left there is no additional effect on the RNA other than levels continuing to rise. This is the point at which there are no longer enough free spliceosomes to sustain the first step of splicing.

A splicing check point failure may lead to transcription reduction, but that does not seem to be the case here as the levels of pre-mRNA rise when splicing is clearly affected. Either transcription cannot be completely blocked or the loss of one of these helicases removes the checkpoint entirely.

The slight dip in pre-mRNA levels at time point 4 (Figure 1b), 36 % Prp16p remaining (Figure 1c), could be due to the absence of Prp16p affecting the step 1 checkpoint. If the spliceosome were to stall at step 1, by whatever cause, Prp16p would trigger a checkpoint and the process leading to spliceosome release by Prp43p commenced, leading to eventual pre-mRNA degradation. If this checkpoint is mediated by Prp16p, which is now absent, there is no checkpoint to fail. The pre-mRNA now has additional time to splice. This checkpoint's absence could be revealed by this observed dip of pre-mRNA levels as some pre-mRNA is successfully spliced rather than waiting to be degraded.

7.3.5 Mild depletion

Prp22p depletion was performed in the milder AID system using the weaker ADH1 promoter to drive TIR1 expression (see chapter 6). This results in slower and lesser depletion of the target protein than the strong ADH1 promoter or the β-est AID system. Again, the depleting strain was thiolabelled (P.Figure 5b, note that the protein was Prp22p not Prp16p as stated at one point in the caption). Under these milder depletion conditions, the depleted Prp22p levels reduced the efficiency of the release of the lariat, shown by their build up. The RNA detected by the branch site PCR does not rise throughout the time course, indicating that step one is not affected. The 3'SS PCR reaction detects a combination of pre-mRNA and lariat-exon2, it's increase is possibly due to the lariat-exon2 product increasing as step 2 fails. Step one does not seem to be affected. It is probable that the cell has accommodated the slower depletion and is releasing stalled spliceosomes at a rate sufficient to maintain first step of splicing even at the lowest levels of Prp22p achieved in this strain.

7.3.6 Global Splicing in a Prp16p Defective Strain

The prp16-2 strain (Vijayraghavan et al., 1989), (Arenas and Abelson, 1997), carries a mutation of the *PRP16* gene conferring temperature sensitivity. The splicing of pp16-2 strain was monitored by microarray (see introductory chapter), on shift from permissive to non-permissive temperatures. Even at permissive temperatures the number of lariats is more than twice as abundant in the prp16-2 strain, compared to the wild-type (Figure 2) indicating a mild problem with second step. Only after shift to non-permissive temperatures do the levels of pre-mRNA begin to rise and exceed the number of lariats. The numbers of lariats accumulating do not reach those of pre-mRNA at later time points indicating that either the cell quickly runs out of spliceosomes. This result is entirely consistent with the RT-qPCR data presented above and shows that failure to recycle spliceosome components has global repercussions on the transcriptome.

7.4 Reinterpretation

This is a recent publication and it has not yet been cited. I have provided additional insights and new interpretations of some of the data in this paper.

7.4.1 Future Work

The global analysis should be repeated with modern techniques, such as RNA-seq (lariat-seq), so accurate lariat levels can be obtained globally. This would show whether different genes have different spliceosome pools, if co- versus post-transcriptional splicing are affected differently by recycling defects and provide an estimate of the spare capacity of spliceosomes in a cell. This analysis would also show if the helicases were no longer able to proofread splicing, by an increase in the use of non-annotated splice sites. Effects on transcription could also be determined.

7.5 Contribution

As discussed in the previous chapter, development of the β-est AID system was a collaborative project in which I took a major part. Specifically, in this paper P.Figure 1 uses the RT-qPCR systems developed by me as does

P.Figure 3. The Prp16 and Prp22 AID strains were made by me and P.Figure 4c is my data. P.Figure 5 is my data using Isabella Maudlin's strain.

All additional data presented in this chapter are my own, as is any re-interpretation of the results and conclusions from the paper.

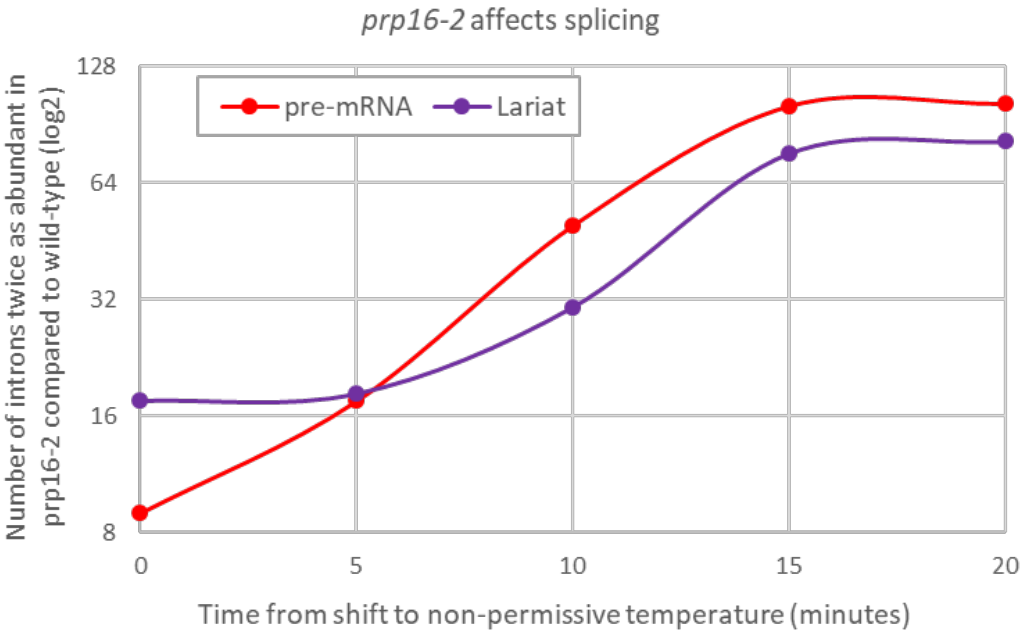


Figure 2: Global analysis of Prp16p mutant splicing
The strain was shifted from permissive temperature (25 °C), to non-permissive (37 °C), in YPDA and samples taken at 5 minute intervals for 20 minutes. cDNA from an isogenic wild-type strain treated as the mutant strain was used as the control. The number of lariats (purple) and pre-mRNAs (red) twice as abundant in the mutant prp16-2 strain compared against the wild-type, minus those that were twice as abundant in the wild-type compared against the prp16-2 strain.
The fluorescent cDNA was made in two ways; using the normal exon reverse transcription primer to assay pre-mRNA and a primer spanning the lariat junction to produce labelled cDNA specifically from lariats (see thesis introduction)
Lariats accumulate even at permissive temperatures, but there is a limit to how many can accumulate before no more can be made as after 5 minutes spliceosome stocks are depleted, first step fails and pre-mRNAs increasingly detected in prp16-2 compared to the wild-type.

7.6 References

Most references are in the publication, a few additional ones are listed here.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc.Natl.Acad.Sci.USA *94*, 11798–11802.

Barrass, J.D., and Beggs, J.D. (2019). Extremely Rapid and Specific Metabolic Labelling of RNA In Vivo with 4-Thiouracil (Ers4tU). JoVE J. Vis. Exp. e59952.

Su, Y.-L., Chen, H.-C., Tsai, R.-T., Lin, P.-C., and Cheng, S.-C. (2018). Cwc23 is a component of the NTR complex and functions to stabilize Ntr1 and facilitate disassembly of spliceosome intermediates. Nucleic Acids Res. *46*, 3764–3773.

Toroney, R., Nielsen, K.H., and Staley, J.P. (2019). Termination of pre-mRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. Genes Dev. *33*, 1555–1574.

Vijayraghavan, U., Company, M., and Abelson, J. (1989). Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. *3*, 1206–1216.

7.7 Reprint

RNA BIOLOGY
2019, VOL. 16, NO. 12, 1775–1784
https://doi.org/10.1080/15476286.2019.1657788

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Blocking late stages of splicing quickly limits pre-spliceosome assembly in vivo

Gonzalo I. Mendoza-Ochoa[a*], J. David Barrass[a], Isabella E. Maudlin[a‡], and Jean D. Beggs[a]

[a]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

ABSTRACT

Pre-messenger RNA splicing involves multi-step assembly of the large spliceosome complexes that catalyse the two consecutive trans-esterification reactions, resulting in intron removal. There is evidence that proof-reading mechanisms monitor the fidelity of this complex process. Transcripts that fail these fidelity tests are thought to be directed to degradation pathways, permitting the splicing factors to be recycled. While studying the roles of splicing factors in vivo, in budding yeast, we performed targeted depletion of individual proteins, and analysed the effect on co-transcriptional spliceosome assembly and splicing efficiency. Unexpectedly, depleting factors such as Prp16 or Prp22, that are known to function at the second catalytic step or later in the splicing pathway, resulted in a defect in the first step of splicing, and accumulation of arrested spliceosomes. Through a kinetic analysis of newly synthesized RNA, we observed that a second step splicing defect (the primary defect) was rapidly followed by the first step of splicing defect. Our results show that knocking down a splicing factor can quickly lead to a recycling defect with splicing factors sequestered in stalled complexes, thereby limiting new rounds of splicing. We demonstrate that this 'feed-back' effect can be minimized by depleting the target protein more gradually or only partially, allowing a better separation between primary and secondary effects. Our findings indicate that splicing surveillance mechanisms may not always cope with spliceosome assembly defects, and suggest that work involving knock-down of splicing factors or components of other large complexes should be carefully monitored to avoid potentially misleading conclusions.

ARTICLE HISTORY
Received 3 June 2019
Revised 3 August 2019
Accepted 14 August 2019

KEYWORDS
Auxin; pre-mRNA splicing; Prp22; protein depletion; yeast

Introduction

Pre-messenger RNA (pre-mRNA) splicing is the process by which introns are removed from RNA transcripts and the coding sequences are joined by two consecutive trans-esterification reactions catalysed by the spliceosome (reviewed in [1–3]). The spliceosome is a multi-megadalton RNA-protein complex that is assembled from five small nuclear ribonuclear protein particles (snRNPs) (U1, U2, U4, U5 and U6 snRNPs) plus non-snRNP proteins, including the nineteen complex (NTC) and NTC-related proteins. Spliceosome assembly is a highly dynamic process (Fig. 1a). Briefly, U1 and U2 snRNPs recognize and bind at the intron 5′ splice site (5′ss) and branch site (BS), respectively, to form the pre-spliceosome, or A complex. Association of the U4/U6.U5 triple snRNP produces a transient pre-B complex from which the U1 snRNP is displaced to produce the more stable B complex, then removal of the U4 snRNP and recruitment of NTC forms the Bact complex, followed by a further reorganization to create the catalytically active B* complex. The first catalytic step of splicing takes place, and the resulting C complex is remodelled again (C* complex), leading to the second catalytic step. Finally, the post-catalytic spliceosome is actively disassembled and the components are recycled.

Thanks to extensive biochemical and genetic studies (reviewed in [4]), together with high-resolution structures obtained by cryo-electron microscopy [5–13], we now have a comprehensive mechanistic understanding of splicing. However, until relatively recently, pre-mRNA splicing was studied mainly as an isolated process whereas, within the context of the cell, splicing functionally interacts with other cellular systems such as transcription, chromatin and RNA processing (reviewed in [14]).

Most splicing factors are essential for viability. Therefore, in vivo studies of the roles of pre-mRNA splicing factors have generally involved the use of conditional mutants (in yeast) or targeted knock-down of individual factors (e.g. by RNAi in higher eukaryotes). In this study, we were particularly interested in Prp16 and Prp22, which are members of the family of DEAH-box RNA-stimulated ATPases, or RNA helicases, that promote structural rearrangements in splicing complexes (reviewed in [15]). Prp16 binds at, or near, the 3′ss, and triggers the formation of C complex and activation of the catalytic core [5,6,16], whereas Prp22 sits downstream of the 3′ss and releases the spliced mRNA from the post-spliceosome [5,17–19]. Prp22 has also been implicated in the second catalytic step of splicing [20] and in 3′ss selection [21,22].

Our goal was to study the effects on splicing efficiency and co-transcriptional spliceosome assembly of knocking down Prp16 or Prp22 in vivo. For comparison, we studied two other splicing

CONTACT Gonzalo I. Mendoza-Ochoa mendoza.ochoa.gi@gmail.com Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3BF, UK
*Current address: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK
‡Current address: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK
Supplementary data for this article can be accessed here.

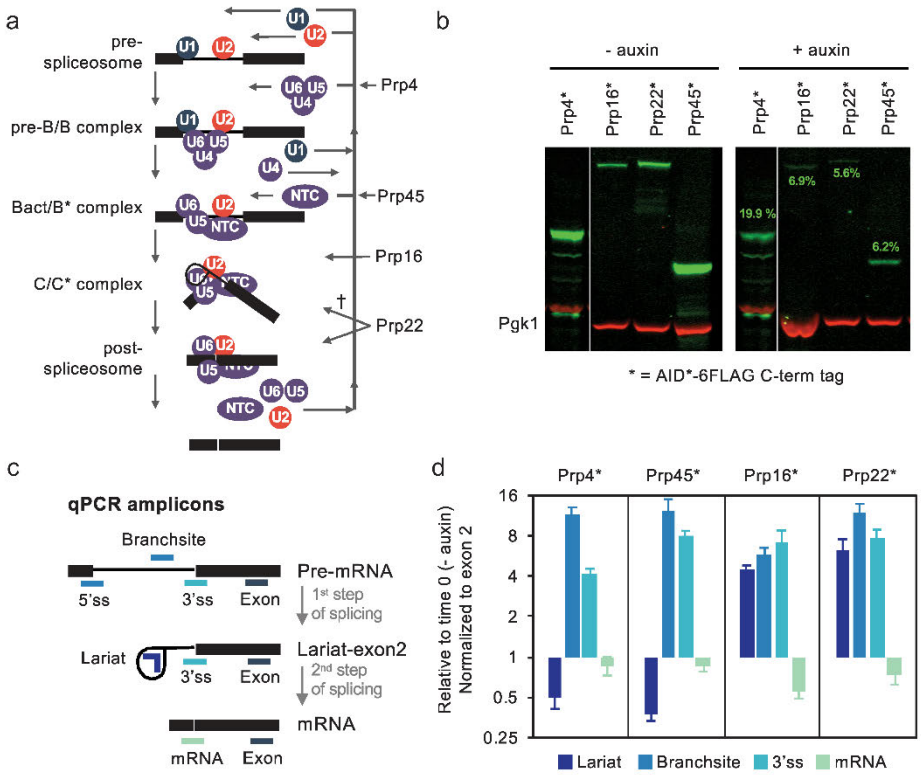


Figure 1. Depletion of splicing factors Prp4, Prp22, Prp16 or Prp45 leads to a first step of splicing defect. (a) Spliceosome assembly model showing recruitment of U1, U2, U4, U5 and U6 snRNPs, nineteen complex (NTC) and splicing factors that were depleted in this study, and their critical stage of activity. (b) Auxin-dependent targeted depletion assessed by immunoblotting with anti-FLAG and anti-Pgk1 (internal control) antibodies. Prp4 was depleted in strain PADH1-409-TIR1. Prp16, Prp22 and Prp45 were depleted in strain PADH1-701-TIR1. Target protein levels after depletion (+ auxin) are shown in green numbers as a percentage of the starting amount, and represent the average of three biological replicates. The anti-FLAG signal was normalized against Pgk1, which is encoded by an intronless gene and was used as internal control. Only one representative blot is shown. (c) Illustration of qPCR-amplified regions from *ACT1* and *ECM33* transcript: 5′ splice site (5′ss) (*ECM33* only), Branchsite (BS) (*ACT1* only), 3′ splice site (3′ss), exon 2, lariat and spliced mRNA. 5′ss and BS primers only detect pre-mRNAs, 3′ss primers will detect pre-mRNA and also lariat-exon2, a product of the first step of splicing. Lariat will detect both the lariat-exon2 and excised lariat (a product of the second step of splicing). mRNA is produced in the second step of splicing. Exon primers are used as controls to normalize for transcription. (d) qPCR of splicing intermediates of *ACT1* at 30 min depletion, normalized to exon 2 and relative to no depletion (time 0). Pre-mRNA accumulation (increase 3′ss and BS) is indicative of a first step of splicing defect. Error bars denote standard error of biological triplicates. † a second step function of Prp22 may not be required for splicing of all intron-containing transcripts[20].

factors, tri-snRNP protein Prp4 and NTC-related protein Prp45, that are involved in different stages of spliceosome assembly. To achieve a fast and specific depletion of our target proteins we used the auxin-inducible degron system [23,24].

Although surveillance mechanisms have the ability to identify defective splicing complexes and target them for dissociation and recycling of the components, our results show that rapid depletion of a splicing factor in vivo can limit the early steps of spliceosome assembly, indicating that the recycling process is overwhelmed. Consequently, the observed phenotypes do not reflect the primary function of the depleted factor. In the case of Prp22, we demonstrate that a more gradual and less complete depletion strategy allows for separation of the primary and secondary effects. We conclude that the budding yeast surveillance and recycling processes cannot cope with large-scale inhibition of splicing, highlighting the need for caution in interpreting the results of in vivo knockdown studies to analyse the roles of different components of a biochemical pathway.

Methods

Yeast strains and growth conditions

See Table S1 for yeast strain genotypes. The OsTIR1 auxin-binding protein was expressed in *S. cerevisiae* strains PADH1-701-TIR1 or PADH1-409-TIR1 (depletes more gradually than PADH-701) directed by constitutive P_{ADH1} promoters, whereas in PZ4EV-NTIR1 it is subject to regulated expression from a β-estradiol-inducible promoter [25] as previously described [23]. Target proteins were C-terminally tagged with AID*-6FLAG (referred to in the text simply as 'AID-tagged') [26], using

a PCR-based method to alter the coding sequence on the genome [27]. Prp16, Prp22 and Prp45 were individually depleted in PADH1-701-TIR1 while Prp4 was depleted in PADH1-409-TIR1 (Figs. 1–3). Depletion of Prp16 in Fig. 4c was done in strain PZ4EV-NTIR1. Where specified, more gradual depletion of Prp22 (Fig. 5) was performed in strain PADH1-409-TIR1. Yeast strains were grown at 30°C on Yeast Peptone Dextrose supplemented with adenine (YPDA). Protein depletions were performed as previously described [23].

Antibodies used

Western blots were performed as previously described [23], using rat anti-FLAG (Agilent, Cat. No. 200474) and mouse anti-PGK1 (Abcam, Cat. No. Ab113687) antibodies. The anti-FLAG signal was normalized against Pgk1, which is encoded by an intronless gene and was used as internal control. Rabbit anti-Prp40 polyclonal antibodies (our laboratory) were used for ChIP of U1 snRNP, and mouse anti-HA 12CA5 monoclonal antibodies (Roche, Cat. No. 11583816001) were used for ChIP and RIP of U2 snRNP.

RNA analysis and chromatin immunoprecipitation

Reverse transcription (RT) was performed as previously described [28]. 2x SYBR green III master mix (Agilent Cat. 600882-51) was used for quantitative polymerase chain reaction (qPCR). Oligonucleotide primers for RT and qPCR are listed in Supplemental Methods. RNA immunoprecipitation (RIP) was performed as previously described [29]. Protocol for chromatin immunoprecipitation-quantitative PCR (ChIP-qPCR) is described in Supplemental Methods. Isolation of total RNA and newly synthesized RNA (nsRNA) by 4-thiouracil (4tU) labelling were performed as described previously [30] with a few modifications [31].

Results

Depletion of Prp16 or Prp22 causes a first step of splicing defect

We investigated the effect on splicing efficiency in vivo of depleting Prp4, Prp16, Prp22 or Prp45. For this we individually C-terminally tagged these proteins with the AID*-6FLAG tag [26] in *Saccharomyces cerevisiae* strains that constitutively express the plant auxin-binding receptor TIR1. We added auxin to cultures of the AID-tagged strains and took samples for analysis immediately (T0) and after 30-min incubation (T30). By western blot analysis, we estimated that Prp4 was depleted to around 20%, and Prp45, Prp16 and Prp22 were depleted to less than 7% of initial values (Fig. 1b). Having confirmed that targeted depletion was successful, we measured the relative abundance of the pre-mRNA, the lariat-exon2 splicing intermediate and spliced mRNA of *ACT1* transcripts by reverse transcriptase real-time quantitative PCR (RT-qPCR) using specific primers (Fig. 1c,d). As anticipated, depletion of tri-snRNP protein Prp4 or NTC-related Prp45 (proteins recruited before the first step of splicing) led to an increase in signal across both the branchsite (BS), representing unspliced pre-mRNA; and 3′ splice site (3'ss), due to unspliced pre-mRNA or lariat-exon2 levels, while lariat levels decreased due to reduced production of lariat-exon2 and/or excised intron. These observations indicate a defect in the first step of splicing. Depletion of Prp16 or Prp22 led to increased lariat and 3'ss signals, indicative of a second step defect but, unexpectedly, also to increased BS levels, suggesting that both steps of splicing were negatively affected by reduction in these late-acting splicing factors. Because it is unlikely that both Prp16 and Prp22 have an additional and uncharacterized role in the first catalytic step of splicing, we speculated that the observed phenotype was an indirect consequence of depleting these proteins.

Pre-spliceosome formation is reduced in the absence of Prp4, Prp45, Prp16 or Prp22

It was previously demonstrated that the co-transcriptional recruitment of splicing factors can be monitored in vivo using the chromatin immunoprecipitation (ChIP) approach, because splicing factors bound to nascent transcripts are sufficiently close to RNA polymerase to interact, either directly or indirectly, with the DNA template [29,32–36]. Therefore, to study the effect on early stages of co-transcriptional spliceosome assembly of depleting Prp16 or Prp22, we performed ChIP of Prp40 and HA-tagged Lea1, as core components of U1 and U2 snRNP, respectively, at the well characterized intron-containing *ACT1* and *ECM33* genes. In the undepleted controls, we observe the typical profiles for ChIP of U1 and U2 snRNP components, with both signals high over exon 2 but U1 snRNP signal declining more 5′ than the U2 signal, representing release of the U1 snRNP as B complex forms (Fig. 2) [33,34]. Following depletion of Prp4, we observe lower occupancy of Prp40 (U1) and Lea1 (U2), compared to the undepleted control. Given that Prp4 joins after pre-spliceosome formation (Fig. 1a), these data suggest that depleting Prp4 also causes an unexpected defect in U1 and U2 snRNP recruitment. In the absence of Prp45, we observe lower occupancy of Lea1 (U2), and higher occupancy of Prp40 (U1) towards the 3′ end of the genes (Fig. 2), indicating reduced and possibly delayed U2 recruitment, which agrees with our previous observation of Prp45 depletion causing a first step of splicing defect. The elevated U1 signal is likely caused by failure to form B complex, which normally involves U1 release [37]. In the absence of either Prp16 or Prp22, we again observe lower occupancy of Lea1 (U2) and higher occupancy of Prp40 (U1), similar to depletion of Prp45. Given that each of these four factors is thought to function after U1 and U2 snRNP recruitment, the reduced ChIP signal for U2 snRNP suggests that depleting any one of these proteins also causes a defect at an earlier than expected stage of spliceosome assembly.

To explain these unusual observations, we hypothesize that depletion of any of these four splicing factors causes accumulation of arrested complexes, with splicing components becoming sequestered, which leads to reduced spliceosome assembly on newly synthesized transcripts and a first step splicing defect.

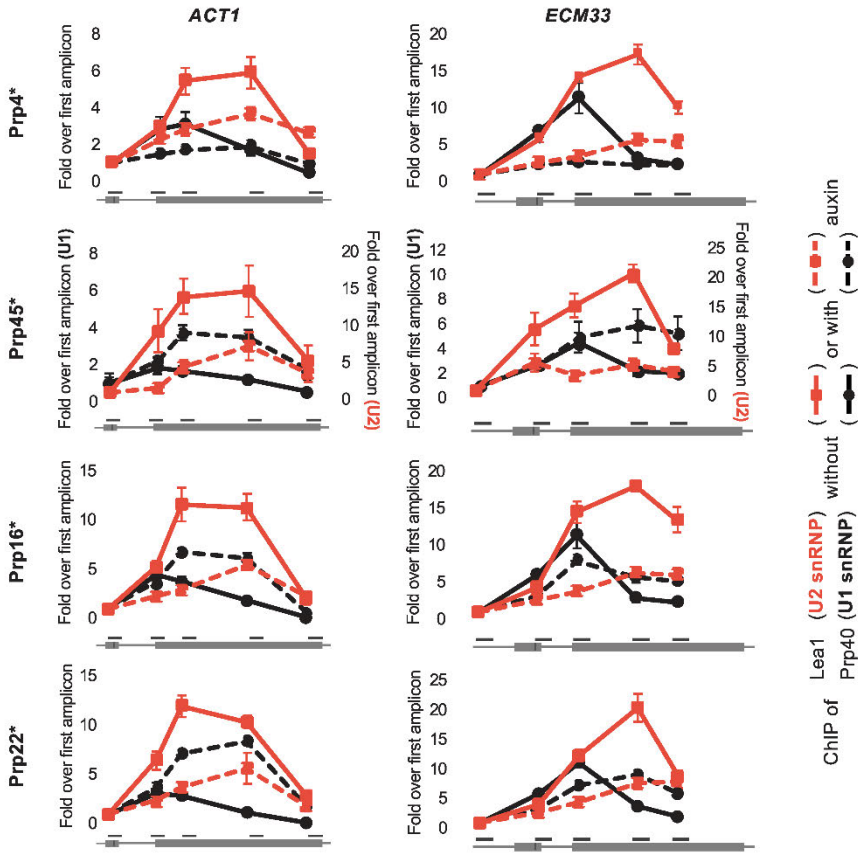


Figure 2. Depletion of several splicing factors (as in Fig. 1) leads to reduced co-transcriptional recruitment of Lea1 (U2 snRNP). ChIP of Prp40 (U1 snRNP; black lines) and HA-tagged Lea1 (U2 snRNP; red lines) on *ACT1* (left plots) and *ECM33* (right plots) genes before (solid lines) and after (dashed lines) auxin-induced depletion. The x-axis represents amplicon location within the gene. For each protein, the ChIP data are presented as relative to the first amplicon for that protein – in the exon 1 of *ACT1* and 5′UTR of *ECM33*. Error bars denote standard error of biological triplicates. U1 and U2 ChIP values following Prp45 depletion are at different scales (y-axis values of U1 ChIP are shown on the left of the graph and U2 ChIP values on the right) due to experimental variations (e.g. variations in immunoprecipitation efficiencies). * = AID*-6FLAG C-terminus tag.

Depletion of splicing factors correlates with increased snRNP interactions

To test the hypothesis that splicing complexes accumulate in cells depleted of these splicing factors, we performed an RNA immunoprecipitation (RIP) analysis where HA-tagged Lea1 (U2) was pulled down and the associated snRNAs were measured by RT-qPCR. The normal pattern of Lea1 (U2 snRNP) association with U1, U4, U5 and U6 snRNAs is presented in Fig. 3a. Prp4 is required for tri-snRNP recruitment, a step necessary for the transition of pre-spliceosome to pre-B complex (Fig. 1a) [38]. Therefore, based on the ChIP data we anticipate that absence of this protein may lead to the accumulation of pre-spliceosome complexes (U1 and U2 snRNPs). Indeed, we observe that depletion of Prp4 correlates with increased association of Lea1 with U1 snRNA, and decreased association with U4, U5 and U6 snRNAs, relative to the undepleted control (Fig. 3b).

In the case of Prp45 depletion, the slightly elevated association of Lea1 with U1 and U4 compared with the undepleted control may indicate accumulation of pre-B and/or B complex, as a result of inefficient conversion of B complex to Bact complex (Fig. 3b). Prp45 is an NTC-related protein, and it has been proposed that the NTC functions at the transition from B to Bact complex by stabilizing the association of the U5 and U6 snRNAs with the 5′ end of the intron after U4 is dissociated [39]. In view of the slightly reduced association of Lea1 with U5, ChIP was performed for the U5 snRNP protein, Prp8, showing that co-transcriptional recruitment of U5 snRNP was also reduced following depletion of Prp45 (Fig S1), adding further support for failure to form a stable B complex.

In contrast, depletion of Prp16 or Prp22, correlates with increased association of Lea1 with U5 and U6 snRNAs, suggesting accumulation of Bact and/or C complexes that contain U2,

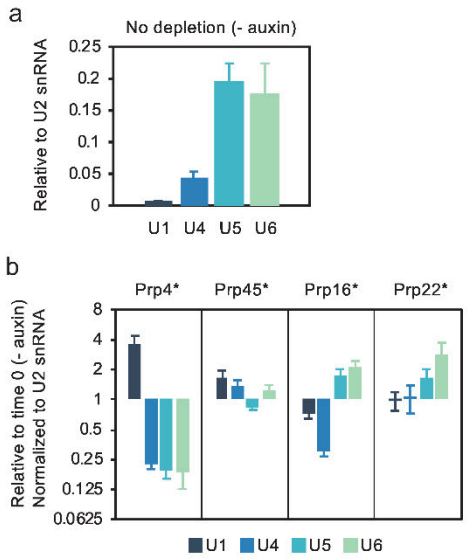


Figure 3. Level of association of U2 snRNP with U1, U4, U5 and U6 snRNAs is altered in the absence of certain splicing factors. (a) RT-qPCR measurement of snRNAs associated with immunoprecipitated, HA-tagged Lea1, a core component of U2 snRNP. (b) RNA immunoprecipitation (RIP) analysis as in panel a after depletion of Prp4, Prp45, Prp16 or Prp22 as in Fig.1. Data are normalized to U2 snRNA signal and presented as relative to no depletion (time 0). Error bars denote standard error of biological triplicates. * = AID*-6FLAG C-terminal tag.

U6 and U5 snRNAs. In the case of Prp16 depletion, there is also decreased association with U1 and U4 snRNAs (Fig. 3b), possibly indicating reduced pre-spliceosome and/or B complex formation. These observations resemble reports of arrested spliceosome accumulation in temperature-sensitive *prp2, prp16* and *prp22* mutant strains [40]. As the ChIP results show reduced co-transcriptional recruitment of U2 snRNP (and likely also of U5 and U4/U6 snRNPs, whose assembly requires pre-spliceosome formation), we conclude that the RIP data reflect the post-transcriptional accumulation of stalled spliceosomes. This can explain both the reduced U2 recruitment to newly synthesized transcripts (Fig. 2) and pre-mRNA accumulation (Fig. 1d).

Kinetic analysis of Prp22 and Prp16 depletion

Next, we studied the kinetics of the splicing defect caused by Prp22 depletion by analysing both 4-thiouracil (4tU)-labelled nascent RNA [30] and co-transcriptional recruitment of Lea1 (U2 snRNP) at short times (0, 3, 6 and 12 min) after auxin addition. Labelling RNA with 4tU in vivo for as little as 1 min allows the isolation of newly synthesized RNA (nsRNA), and its production and processing can be assessed as the protein is depleted, whereas total RNA (Fig. 1) includes RNA produced prior to target protein depletion. The 4tU splicing analysis shows *ACT1* lariat abundance elevated compared to the signal from exon 2 in the sample incubated with auxin for 3 min, then decreasing (relative to exon 2), as the BS (pre-mRNA) signal builds up (Fig. 4a). Moreover, the 3'ss signal also accumulates transiently in parallel with the lariat signal, levels off as lariat declines and BS increases, then accumulates again in parallel with the BS signal. These results indicate that the lariat-exon2 product of the first step of splicing accumulates transiently before a first step defect kicks in and prevents further lariat-exon2 production.

Consistently, the ChIP signal for Lea1 (U2 snRNP) at *ACT1*, is reduced in the 6-min sample but not in the 3-min sample (Fig. 4b), indicating reduced recruitment of U2 snRNP to nascent RNA after only 6 min of Prp22 depletion. This can explain the first step defect detected at 6 min as being due to reduced co-transcriptional spliceosome assembly. Taken together, the kinetic analyses indicate that although the first step splicing defect occurs extremely rapidly after auxin addition, it is preceded by a second step defect, arguing that the first step splicing defect is not a direct consequence of Prp22 depletion, but a secondary effect. In the case of *ECM33* transcript, we observe 5'ss (pre-mRNA) signal increase at 6 min, and also reduced recruitment of U2 snRNP (Lea1) already at 3 min (Fig. 4a, B; specific assay of *ECM33* lariat by RT-qPCR has not been achieved).

We performed a similar kinetic analysis of Prp16 depletion, this time using a Prp16 AID-tagged strain in which TIR1 is conditionally expressed by the addition of β-estradiol prior to auxin addition, permitting tighter control of the system [23]. The results were similar to those for Prp22 depletion, with lariat signal peaking at 6 min after auxin addition, then decreasing as the BS signal increased but, notably, after a significant lag (Fig. 4c).

A more gradual depletion of Prp22 increases contrast between primary and secondary effects

As depletion of Prp22 in the PADH1-701-TIR1 strain affected splicing extremely quickly, and from a situation where this protein was already partially depleted without auxin, we analysed another AID strain, PADH1-409-TIR1, that expresses TIR1 at a lower level and, therefore, allows a more gradual depletion of Prp22 and from a starting point of no auxin-independent depletion [23]. Analysis of total RNA (Fig. 5a) showed that reduction of Prp22 to 42% of the initial level took 15 min as opposed to 3 min with the more rapidly depleting strain (shown in Fig. 4b), and caused only a low level of lariat accumulation. At 30 min, when Prp22 was reduced to 20%, there was accumulation of more lariat, as well some 3'ss signal and, to a lesser extent, BS signal, indicating a small amount of pre-mRNA. At 60 min, with only 9% Prp22 remaining, the signals for lariat, 3'ss and BS were all strongly elevated.

4tU-labelling analysis of the kinetics of splicing with gradually depleted Prp22 (Fig. 5b) also showed that 15 min after auxin addition, with 42% of Prp22 remaining, there was significant accumulation of only the lariat signal, indicative of excised intron lariat accumulation. At 30 min (Prp22 at 20%) 3'ss signal remained low, whereas at 60 min (Prp22 at 9%) the 3'ss signal accumulated at a similar rate to lariat signal, and there was very little BS (pre-mRNA) accumulation, indicating build-up of lariat-exon2. In summary, during a more gradual Prp22 depletion, excised intron lariat clearly accumulated before the lariat-exon2 product of the first step, and there was little pre-mRNA

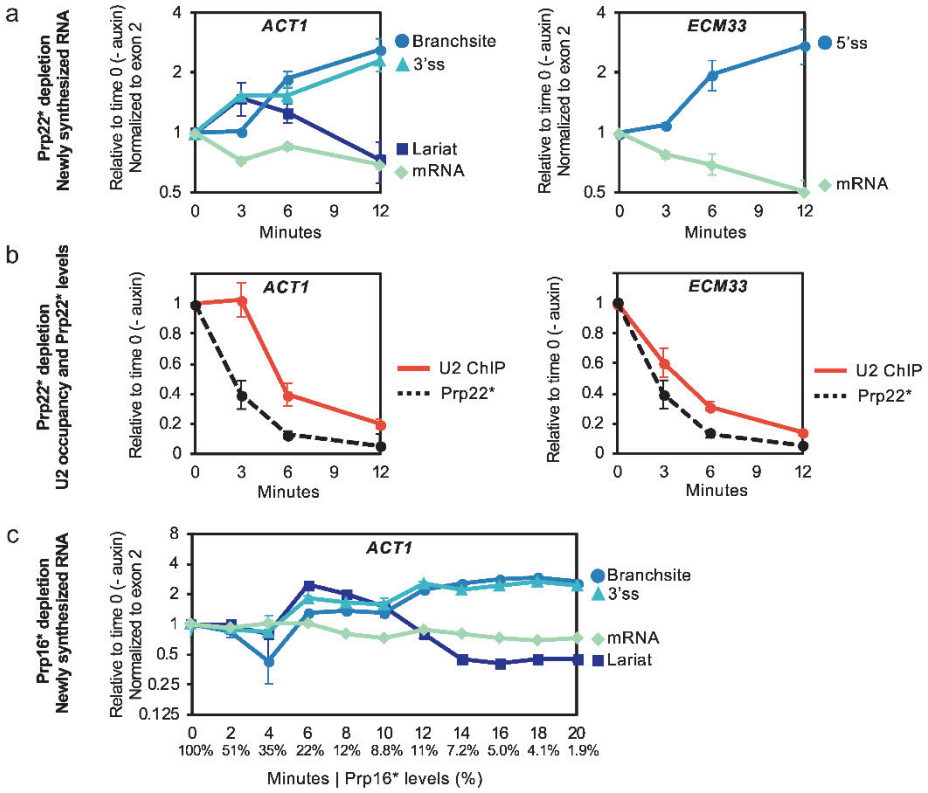


Figure 4. Kinetic analysis of newly synthesized RNA after depletion of Prp22 or Prp16. (a) RT-qPCR analysis of 4tU-labelled newly synthesized RNA of *ACT1* and *ECM33*. Following Prp22 depletion (in strain PADH1-701-TIR1) for the times indicated, samples were 4tU-labelled for 1 min. Data are normalized to conditions without depletion. (b) Occupancy of U2 snRNP-core component Lea1 was measured from depletion time course samples as in panel a. Solid red lines are Lea1 ChIP-qPCR for *ACT1* and *ECM33*, and dashed black lines are Prp22 levels measured by western blot (blot image not shown). ChIP data are normalized against background (intronless *ALG9*). Prp22 western blot signal is normalized against internal control Pgk1. (c) RT-qPCR analysis of 4tU-labelled (1-min labelling) newly synthesized *ACT1* RNA, following Prp16 depletion in strain PZ4EV-NTIR1, with osTIR1 expression induced by β-estradiol[23]. Relative levels (%) of Prp16 protein are in X-axis below minutes. Data in panels (a) and (c) are normalized against exon 2. Data in all panels (a-c) are relative to no depletion (time 0) and x-axes represent time (minutes) after addition of auxin. Error bars denote standard error of four (panels a-b) or three (panel c) biological replicates.

accumulation, indicating that, under these milder depletion conditions, the first step of splicing was hardly affected.

Discussion

Our observation of pre-mRNA accumulation when depleting Prp16 or Prp22 resembles previous studies showing pre-mRNA accumulation at the restrictive temperature in strains with heat-sensitive mutants of *PRP16* [41,42] and *PRP22* [17,41–43]. A homozygous lethal mutation of the zebrafish orthologue of Prp22 (called Dhx8), also caused accumulation of unspliced pre-mRNA [44], indicating that this phenotype is not specific to yeast. Company et al. [43]speculated that in a heat-sensitive *prp22* strain unspliced pre-mRNA accumulated because defective spliceosomes were not recycled for new rounds of splicing, but this was not investigated further. Evidence supporting the concept of defective spliceosomes limiting an earlier step came subsequently, in a report that metabolic depletion of U5 snRNA caused accumulation of arrested pre-spliceosomes and reduced co-transcriptional recruitment of U1 and U2 snRNPs [29], which is similar to our Prp4 depletion analysis. However, as the *GAL* promoter-driven expression of U5 snRNA was repressed for 16 h, it cannot be ruled out that the recycling defect was an indirect consequence of the prolonged splicing defect. For example, reduced expression of intron-containing genes that encode splicing factors (e.g. U1 snRNP protein Mud1) could explain the reduced co-transcriptional recruitment of U1 snRNP that was observed.

Our AID depletion approach is novel because (1) we target not just one, but different stages of splicing, (2) we deplete the target proteins very rapidly, (3) we analyse both the efficiency of splicing newly synthesized transcripts and co-transcriptional snRNP recruitment, and (4) we demonstrate the accumulation of different splicing complexes depending on the stage at which the depleted factor functions. Taken together, data derived from this approach and from previous studies [29,40] demonstrate that perturbing splicing can lead to a recycling defect that limits the earliest steps of the splicing cycle. We show that this 'feed-

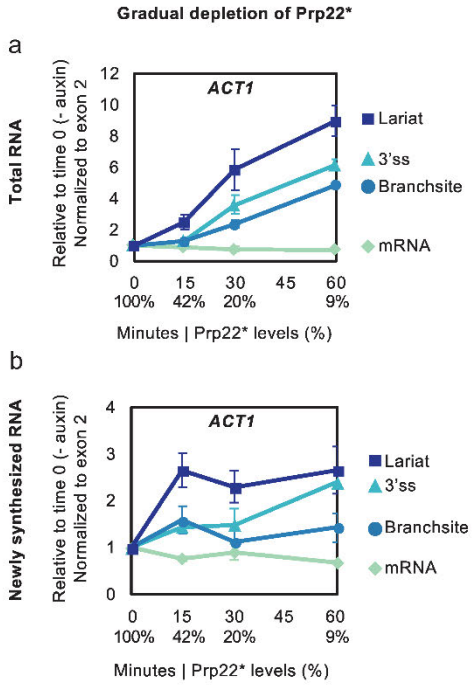


Figure 5. RNA Analysis in a time course of a more gradual Prp22 depletion. Prp22 was depleted in strain PADH1-409-TIR1, in which OsTIR1 is expressed to low levels[23] so the Prp22 level declines neither as far nor as fast as in the other figures. Splicing intermediates of *ACT1* were analysed by RT-qPCR from total RNA (a) and 4tU-labelled (labelled for 4 min at each time of depletion) newly synthesized RNA (b). Data are normalized to exon 2 and relative to no depletion (time 0). Relative levels (%) of Prp16 protein are in X-axis below minutes. Error bars denote standard error of biological triplicates.

back' effect happens extremely rapidly, suggesting that spliceosome disassembly can become a rate-limiting step for splicing, and supporting the proposal that pre-mRNA substrates compete for a limited pool of spliceosome components in yeast [45]. Moreover, we demonstrate that the effect is indeed titratable; partial and/or slower depletion of Prp22 resulted in a less severe defect, allowing the kinetics of this effect to be analysed. Therefore, we propose that by controlling the rate and optimizing the extent of depletion, one can minimize secondary effects associated with recycling defects and thereby facilitate the functional analysis of a targeted protein.

Similar to our results for Prp45 depletion, Hálová et al. [46] also observed a reduction in co-transcriptional U2 snRNP recruitment when they analysed the effect of truncating the C-terminal domain of Prp45, proposing that Prp45 has an earlier than expected role in splicing, at the stage of U2 snRNP recruitment. Our observation that Prp45 depletion leads to enhanced U1 snRNP and reduced U2 snRNP signal at the site of transcription could support either a role for Prp45 in U2 snRNP recruitment or, alternatively, may indicate a recycling defect. Prp45 depletion also resulted in a mildly increased association of U2 with U1 and U4 (our RIP analysis), possibly indicating pre-B and/or B complex accumulation (Fig. 3b). This seems plausible, as Prp45 is an NTC-related protein, and it has been proposed that the NTC functions at the transition from B to Bact complex by stabilizing the association of the U5 and U6 snRNAs with the 5' end of the intron after U4 is dissociated [39]. Prp45 has an extended conformation in the spliceosome [12], interacting with many spliceosome components. It is conceivable that depletion of Prp45 could directly affect more than one step in the splicing cycle, perhaps even in a substrate-specific manner, which could make mutation or depletion results difficult to interpret.

Chen et al. [47] tested the ability of the NTR disassembly complex (containing Prp43, Ntr1 and Ntr2) to mediate substrate release from the spliceosome when various DEAH-box RNA-stimulated ATPases were mutated or immunodepleted in vitro. They concluded that only spliceosome intermediate complexes that were arrested after, but not before, the ATP-dependent action of Prp2, Prp16 or Prp22 could be actively disassembled. Our in vivo analyses of Prp16 and Prp22 depletion are also consistent with failure to disassemble complexes stalled before these proteins act. Furthermore, our analyses show that recycling defects are not limited to knockdown of RNA-stimulated ATPases, as depletion of Prp4 or Prp3, another tri-snRNP protein, displayed similar defects (Fig. S2).

The DEAD/H-box RNA-stimulated ATPases were proposed to function as proofreading or fidelity factors in splicing based on assays using reporter genes encoding aberrant or suboptimal pre-mRNA substrates (e.g. pre-mRNAs with non-canonical splice sites or branchsites) [48,49]. It was further proposed that defective complexes are removed by a discard pathway [50,51]. In this work, we focused on a different type of aberrant spliceosomes – those that contain normal pre-mRNA substrates but defective splicing machinery. This situation leads to a wide-spread defect, that may overwhelm surveillance pathways, such that the splicing factors become sequestered in defective complexes, limiting the assembly of new spliceosomes. Interestingly, in the strain where Prp22 was more gradually depleted, accumulation of pre-mRNA (Branchsite) was both slower than lariat accumulation in the same conditions (Fig. 5a) and reduced compared to fast depletion of Prp22 (Fig. 1d), suggesting that a recycling defect is more likely when the target is rapidly depleted, and/or knocked down to low levels. Considering all evidence, it seems that the splicing surveillance machinery likely has the capability to recognize aberrant spliceosomes that either lack an essential splicing factor or contain a mutant protein. This notion is supported by evidence that mutations in disassembly factors Spp382 (Ntr1) and Prp43 suppress the defects caused by the spliceosome assembly mutations *prp38-1* and *prp8-1*, presumably by reducing dissociation of the defective complexes, thereby allowing more time for the defective splicing factors to function [52]. However, our results suggest that the splicing surveillance and/or discard machinery is not equipped to deal with a large or rapid accumulation of aberrant spliceosomes. We propose that when the splicing defect is highly penetrant, the surveillance capacity of the cell can be overwhelmed, leading to the build-up of arrested complexes. We speculate that this might also help to explain why some

disease-causing splicing factor alleles in humans are more penetrant than others.

Our observed rapid accumulation of excised lariat upon Prp22 depletion (Fig. 5) is consistent with Prp22's role in releasing spliced mRNA from the post-spliceosome [17–19], a step that ultimately leads to spliceosome disassembly by the NTR complex and recycling of the components. The high level of lariat and 3'ss signals but low level of branchsite signal for *ACT1* transcripts (indicative of a second step of splicing defect) after 3 min of rapid depletion (Fig. 4a) or 60 min of gradual depletion (Fig. 5b) of Prp22 in the 4tU-labelling analyses supports Prp22 having an additional role in the second catalytic step of splicing [20]. However, as excised lariat accumulated prior to lariat-exon2 (Fig. 5) it suggests that, in vivo, Prp22's role in the second step of splicing is secondary to its role in mRNA release. An alternative explanation could be that accumulation of post-splicing complex following Prp22 depletion reduces the availability of second step factors such as Prp18 and Slu7 to below a limiting amount, thereby affecting the second step indirectly. Both Schwer & Gross [20] and our analyses of Prp22's role in splicing were conducted only on *ACT1* transcripts. Therefore, it remains an open question whether this protein is also required for the second step of splicing of other transcripts.

Different intron-containing transcripts may be more or less sensitive to the depletion of different factors, and splicing may be affected at different stages. It is not always possible to distinguish first and second step splicing defects by RT-qPCR, as distinguishing unspliced pre-mRNA and lariat-exon2 intermediate species (as we have done for *ACT1*) generally depends on the specific measurement of lariats, which can be problematic. However, techniques have been developed to simultaneously sequence lariats of potentially all intron-containing transcripts of *S. cerevisiae* [53–55]. In principle, these transcriptome-wide approaches could be used to investigate, for example, whether Prp22 is required in vivo for the second step of splicing of only a subset of introns as has been proposed [20].

Our observations highlight a concern that, in order to avoid misleading conclusions when knocking down components of large complexes, it is important to consider possible indirect effects that can occur as a consequence of the accumulation of a disabled complex. This is a particular concern if the complex assembles and/or functions in a step-wise manner, such as the spliceosome or the ribosome assembly pathway. Ways of controlling for this include performing kinetic analyses of the effects, comparing the effects of depleting the target factor to different extents and at different rates, or of depleting different components of the complex. Therefore, when interpreting the results of large scale knock-down studies, for example, RNAi depletion analyses of human splicing factors [56–58], the possibility of confounding secondary effects should be considered.

Author contributions

GIMO performed the experimental work, except for Fig. 4C (JD Barrass) and Fig. 5 (IEM and JD Barrass). JD Beggs conceived of the project and supervised the work. GIMO and JD Beggs wrote the manuscript.

Acknowledgments

We thank our colleagues Jane A.E. Reid and Emanuela Sani who provided reagents and technical expertise in AID protein-depletion and ChIP techniques, respectively.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a scholarship to GIMO from the Consejo Nacional de Ciencia y Tecnología, Mexico (CONACYT) and the University of Edinburgh School of Biological Sciences, a Wellcome 4-year PhD studentship [105256] to IEM, and by Wellcome funding [104648] to JD Beggs. Work in the Wellcome Centre for Cell Biology is supported by Wellcome core funding [092076].

ORCID

J. David Barrass http://orcid.org/0000-0002-1624-8276
Jean D. Beggs http://orcid.org/0000-0002-8179-6519

References

[1] Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3. Available from: https://cshperspectives.cshlp.org/content/3/7/a003707.long
[2] Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol. 2017;18:655–670. Available from: https://www.nature.com/articles/nrm.2017.86
[3] Fica SM, Nagai K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. Nat Struct Mol Biol. 2017;24:791–799. Available from: https://www.nature.com/articles/nsmb.3463
[4] Mayerle M, Guthrie C. Genetics and biochemistry remain essential in the structural era of the spliceosome. Methods. 2017;125:3–9. Available from: https://www.sciencedirect.com/science/article/pii/S1046202316304297
[5] Fica SM, Oubridge C, Galej WP, et al. Structure of a spliceosome remodelled for exon ligation. Nature. 2017;542:377–380. Available from: http://www.nature.com/doifinder/10.1038/nature21078
[6] Galej WP, Wilkinson ME, Fica SM, et al. Cryo-EM structure of the spliceosome immediately after branching. Nature. 2016;537:197–201. Available from: http://www.nature.com/doifinder/10.1038/nature19316
[7] Hang J, Wan R, Yan C, et al. Structural basis of pre-mRNA splicing. Science. 2015;349:1191–1198. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26292705
[8] Liu S, Li X, Zhang L, et al. Structure of the yeast spliceosomal postcatalytic P complex. Science. 2017;358:1278–1283. Available from: https://science.sciencemag.org/content/358/6368/1278.
[9] Plaschka C, Lin PC, Nagai K. Structure of a pre-catalytic spliceosome. Nature. 2017;546:617–621.
[10] Wan R, Yan C, Bai R, et al. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. Science. 2016;353:895–904. Available from: https://science.sciencemag.org/content/353/6302/895.

[11] Wilkinson ME, Fica SM, Galej WP, et al. Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. Science. 2017;358:1283–1288. Available from: https://science.sciencemag.org/content/358/6368/1283.

[12] Yan C, Hang J, Wan R, et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. Science. 2015;349:1182–1191. Available from: https://science.sciencemag.org/content/349/6253/1182.

[13] Bai R, Wan R, Yan C, et al. Structures of the fully assembled saccharomyces cerevisiae spliceosome before activation. Science. 2018;360:1423–1429. Available from: https://science.sciencemag.org/content/360/6396/1423

[14] Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet. 2014;15:163–175. Available from: https://www.nature.com/articles/nrg3662.

[15] Cordin O, Beggs JD. RNA helicases in splicing. RNA Biol. 2013;10:83–95. Available from: https://www.tandfonline.com/doi/full/10.4161/rna.22547

[16] Tseng C-K, Liu H-L, Cheng S-C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA. 2011;17:145–154. Available from: https://rnajournal.cshlp.org/content/17/1/145.

[17] Wagner JDO, Jankowsky E, Company M, et al. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. Embo J. 1998;17:2926–2937. Available from: http://emboj.embopress.org/content/17/10/2926.

[18] Aronova A, Bacíková D, Crotti LB, et al. Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. RNA. 2007;13:1437–1444. Available from: https://rnajournal.cshlp.org/content/13/9/1437.

[19] Schwer B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell. 2008;30:743–754. Available from: https://www.sciencedirect.com/science/article/pii/S1097276508003316.

[20] Schwer B, Gross C. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. Embo J. 1998;17:2086–2094. Available from: http://emboj.embopress.org/content/17/7/2086.

[21] Mayas RM, Maita H, Staley JP. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol. 2006;13:482–490. Available from: https://www.nature.com/articles/nsmb1093.

[22] Semlow DR, Blanco MR, Walter NG, et al. Spliceosomal DEAH-Box ATPases remodel Pre-mRNA to activate alternative splice sites. Cell. 2016;164:985–998.

[23] Mendoza-Ochoa GI, Barrass JD, T'erlouw BR, et al. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. Yeast. 2018:1–7. Available from: http://doi.wiley.com/10.1002/yea.3362

[24] Nishimura K, Fukagawa T, Takisawa H, et al. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods. 2009;6:917–922. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19915560

[25] McIsaac RS, Oakes BL, Wang X, et al. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Res. 2013;41:e57. Available from: https://academic.oup.com/nar/article/41/4/e57/2414641

[26] Morawska M, Ulrich HD. An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast. 2013;30:341–351. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23836714

[27] Longtine M, McKenzie A, Demartini D, et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast. 1998;951:943–951. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-0061(199807)14:10%3C953::AID-YEA293%3E3.0.CO;2-U

[28] Alexander R, Barrass J, Dichtl B, et al. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in Saccharomyces cerevisiae. RNA. 2010;16:2570–2580. Available from: http://rnajournal.cshlp.org/content/16/12/2570.

[29] T'ardiff D, Rosbash M. Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. RNA. 2006;12:968–979. Available from: http://rnajournal.cshlp.org/content/12/6/968

[30] Barrass JD, Reid JEA, Huang Y, et al. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol. 2015;16:282. Available from: http://genomebiology.com/2015/16/1/282

[31] Barrass JD, Beggs JD. Extremely rapid and specific metabolic labelling of RNA in vivo with 4-thiouracil (ers4tU). J Vis Exp. 2019;150. Available from: https://www.jove.com/video/59952

[32] Kotovic K, Lockshon D, Boric L, et al. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol Cell Biol. 2003;23:5768–5779. Available from: http://mcb.asm.org/content/23/16/5768.

[33] Lacadie SA, Rosbash M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell. 2005;19:65–75. Available from: https://www.sciencedirect.com/science/article/pii/S1097276505013122.

[34] Görnemann J, Kotovic KM, Hujer K, et al. Cotranscriptional Spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell. 2005;19:53–63. Available from: https://www.sciencedirect.com/science/article/pii/S1097276505013134.

[35] Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell. 2010;40:571–581. Available from: https://www.sciencedirect.com/science/article/pii/S1097276510008427.

[36] Brugiolo M, Herzel L, Neugebauer KM. Counting on co-transcriptional splicing. F1000Prime Rep. 2013;5:9. Available from: https://f1000.com/prime/reports/b/5/9.

[37] Price AM, Görnemann J, Guthrie C, et al. An unanticipated early function of DEAD-box ATPase Prp28 during commitment to splicing is modulated by U5 snRNP protein Prp8. RNA. 2013;20:46–60. Available from: https://rnajournal.cshlp.org/content/20/1/46.

[38] Lygerou Z, Christophides G, Séraphin B. A novel genetic screen for snRNP assembly factors in yeast identifies a conserved protein, Sad1p, also required for pre-mRNA splicing. Mol Cell Biol. 1999;19:2008–2020. Available from: https://mcb.asm.org/content/19/3/2008.short.

[39] Chan S-P, Kao D-I, Tsai W-Y, et al. The Prp19p-associated complex in spliceosome activation. Science. 2003;302:279–282. Available from: https://science.sciencemag.org/content/302/5643/279

[40] Lardelli RM, Thompson JX, Yates JR, et al. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA. 2010;16:516–528. Available from: https://rnajournal.cshlp.org/content/16/3/516.

[41] Noble SM, Guthrie C. Identification of novel genes required for yeast Pre-mRNA splicing by means of cold-sensitive mutations. Genetics. 1996;143:67–80. Available from: https://www.genetics.org/content/143/1/67.

[42] Vijayraghavan U, Company M, Abelson J. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. 1989;3:1206–1216. Available from: http://genesdev.cshlp.org/content/3/8/1206.

[43] Company M, Arenas J, Abelson JN. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature. 1991;349:487–493. Available from: https://www.nature.com/nature/journal/v349/n6309/abs/349487a0.html.

[44] English MA, Lei L, Blake T, et al. Incomplete splicing, cell division defects, and hematopoietic blockage in dhx8 mutant zebrafish. Dev Dyn. 2012;241:879–889. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/dvdy.23774.

[45] Munding EM, Shiue L, Katzman S, et al. Competition between Pre-mRNAs for the splicing machinery drives global regulation of splicing. Mol Cell. 2013;51:338–348.

[46] Hálová M, Gahura O, Převorovský M, et al. Nineteen complex-related factor Prp45 is required for the early stages of

cotranscriptional spliceosome assembly. RNA. 2017;23:1512–1524. Available from: http://rnajournal.cshlp.org/content/23/10/1512.

[47] Chen H-C, Tseng C-K, Tsai R-T, et al. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol Cell Biol. 2013;33:514–525. Available from: https://mcb.asm.org/content/33/3/514.

[48] Burgess S, Couto JR, Guthrie C. A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell. 1990;60:705–717. Available from: https://www.sciencedirect.com/science/article/pii/009286749090086T.

[49] Koodathingal P, Staley JP. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. RNA Biol. 2013;10:1073–1079. Available from: https://www.tandfonline.com/doi/full/10.4161/rna.25245.

[50] Burgess SM, Guthrie C. A mechanism to enhance mRNA splicing fidelity: The RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell. 1993;73:1377–1391. Available from: https://www.sciencedirect.com/science/article/pii/009286749390363U.

[51] Konarska MM, Query CC. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. Mol Cell. 2004;14:343–354. Available from: https://www.sciencedirect.com/science/article/pii/S1097276504002175.

[52] Pandit S, Lynn B, Rymond BC. Inhibition of a spliceosome turnover pathway suppresses splicing defects. Proc Natl Acad Sci U S A. 2006;103:13700–13705. Available from: https://www.pnas.org/content/103/37/13700.

[53] Gould GM, Paggi JM, Guo Y, et al. Identification of new branch points and unconventional introns in Saccharomyces cerevisiae. RNA. 2016;22:1522–1534. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27473169.

[54] Stepankiw N, Raghavan M, Fogarty EA, et al. Widespread alternative and aberrant splicing revealed by lariat sequencing. Nucleic Acids Res. 2015;43:8488–8501. Available from: https://academic.oup.com/nar/article/43/17/8488/2414403.

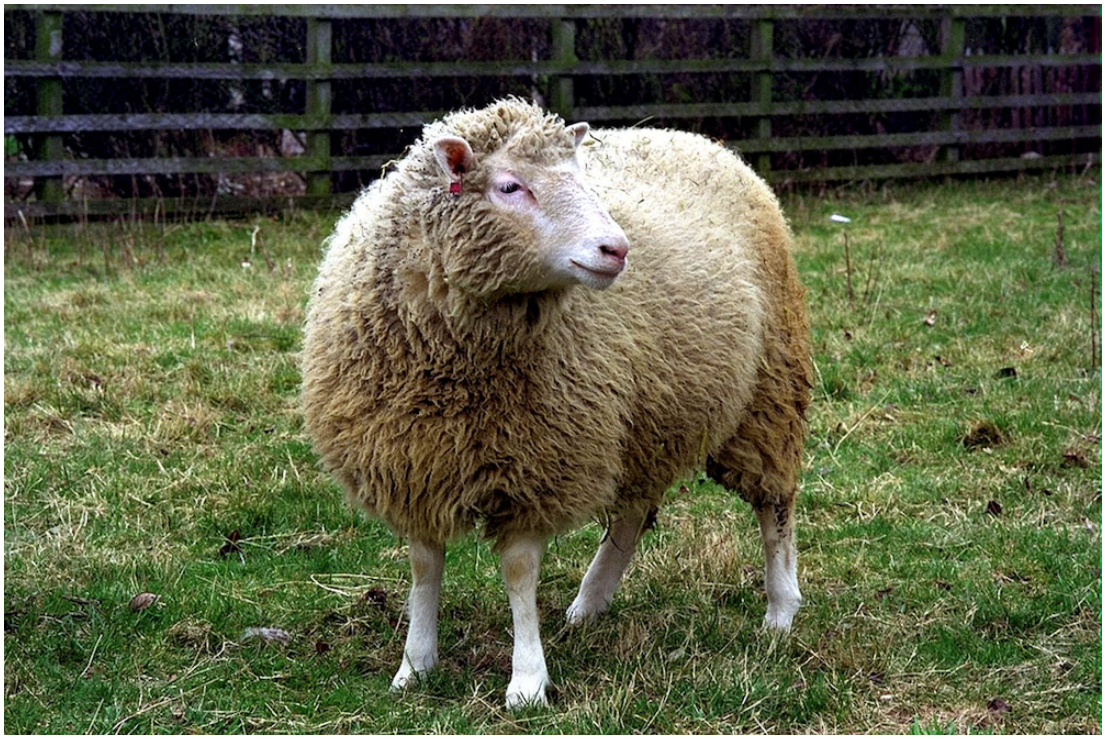
[55] Qin D, Huang L, Wlodaver A, et al. Sequencing of lariat termini in S. cerevisiae reveals 5′ splice sites, branch points, and novel splicing events. RNA. 2016;22:237–253. Available from: https://rnajournal.cshlp.org/content/22/2/237.

[56] Park JW, Parisky K, Celotto AM, et al. Identification of alternative splicing regulators by RNA interference in Drosophila. Proc Natl Acad Sci. 2004;101:15974–15979. Available from: https://www.pnas.org/content/101/45/15974.

[57] Moore MJ, Wang Q, Kennedy CJ, et al. An alternative splicing network links cell-cycle control to apoptosis. Cell. 2010;142:625–636.

[58] Papasaikas P, Tejedor JR, Vigevani L, et al. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. Mol Cell. 2015;57:7–22. Available from: https://www.sciencedirect.com/science/article/pii/S109727651400865X.

8 Concluding Remarks



Dolly the Cloned Sheep

For the entire 1990's, prior to joining the University of Edinburgh, I was employed by PPL therapeutics, eventually rising to Head of Molecular Genetics. It was there I gained experience of the pharmaceutical industry, my concepts of quality control (QC), Good Manufacturing Practice (GMP) and the importance of a well validated protocol. All of these skills have been invaluable since.

The highlight of my time there was without doubt the birth of Dolly and, while I had nothing to do with her creation, I shared an office and lab with those that did. I was also the second person in the world to know that Dolly's telomeres were short as I knew how to operate the Phosphorimager, used in the final step of the assay.

Dolly was very friendly and would come when called. She ate from my hand several times, as she did from anyone's; as a consequence, she was a rather stout sheep. I am sad to see her in a glass case at the National Museum of Scotland but I will always be proud of the small part I played in her life.

8.1 Transcription

As measured by the highly characterised Ribo1 reporter (chapter 2), transcripts can be detected by RT-qPCR less than 5 minutes after addition of the inducing chemical. As a single copy can be detected, this is the time for movement of the chemical into the cell, recruitment and assembly of transcription factors and, finally, for transcription to reach regions of the reporter that can be assayed. As the same study produced a transcription rate of 60 to 90 seconds per kb, transcription can only be a short portion of this time. Minimum labelling times from 4tU studies (chapter 4) of 15 to 30 seconds hint that small molecule import can be very fast; giving a very rough estimation for induction of productive transcription in the region of 4 minutes.

In a population of cells, the transcript copy number is consistent and reproducible; among individual members of that population, however, transcript copy numbers are not constant, and in a single cell stochastic factors predominate. Initiation rates are probably where this variation comes from as many cells do not produce Ribo1 transcript at all.

3' end formation in Ribo1 is also somewhat variable, several cleavage sites can be used, although in 2 tight clusters. Poly-adenylation takes about 30 seconds to complete, as determined by the 3’ end cleavage assay (chapter 2)

8.2 Co-transcriptional Splicing

This cleavage assay (chapter 2) provides incontrovertible evidence of co-transcriptional splicing. After induction, mRNA is detected by reverse transcription from a primer downstream of the cleavage site; the cleavage primer. Detection of spliced products in cDNA generated from this primer can only arise if splicing occurred co-transcriptionally. However, in the first detected transcripts, a similar amount of unspliced pre-mRNA is detected by an oligo dT primer (to the poly(A) tail), indicating that both post- and co-transcriptional splicing occurs initially. However, this is only transient. After the initial burst of transcription very few transcripts reaching the cleavage site are

unspliced. It is possible that the first round of transcription is special; a pioneering round.

The cleavage assay shows 6-8 transcripts arriving at the 3' end in a single sample and not in the sample taken 30 seconds earlier. A possibility is that a train of polymerases forms at the promoter and/or is entrained by a pause of transcription. This pause could be to remove nucleosomes, for factor binding (e.g. transcription elongation and splicing factors), stochasticity during elongation or all three.

The generality of co-transcriptional splicing is reinforced by 4tU metabolic labelling of RNA *in vivo* (chapter 3). This labelling has been optimised to be highly specific for nsRNA, allowing very high temporal resolution without substantial damage to the cells and arguably free of bias. This technique has not provided evidence of a transcriptional pause to allow completion of splicing but cannot rule out a pause for spliceosomal assembly, first step of splicing or to establish splicing on a newly induced transcript.

8.3 Splicing

Thio-labelling of RNA (chapter 4) shows that RPGs are different from other transcripts and almost invariably co-transcriptionally splice. Intron features that inhibit co-transcriptional splicing include large numbers of uracils, 5'SS to BPA secondary structure and particularly the BPA being trapped in an RNA stem. All features generally absent in RPGs.

Splicing has links to other RNA metabolic processes besides transcription, possibly by eIF4G (chapter 5). This protein, a translation factor, has binding sites for cap binding complex, and splicing factors. These are probably mutually exclusive and could be part of the handover of the RNA from cap binding to the splicing machinery and then onto translation. eIF4G depletion only has a minor effect on splicing but could prove vital for splicing if both copies of the gene were removed.

8.4 Quality Control

In other processes where accuracy is important, such as the cell cycle, quality control is maintained by checkpoints that have to be satisfied before the next step can start. There is evidence (chapter 2), for a similar system in splicing. In the Ribo 5'SS reporter (with a deactivated 5'SS), pre-mRNA levels plateau 7.5 minutes post induction. This indicates that transcription has reduced and/or degradation of the transcript has started and now matches the rate of transcription. Indeed, the half-life of this transcript is very short, only about 2 minutes. As U1 snRNPs cannot bind, quality control must involve the proteins, such as Msl5p, that bind to the BS and 3'SS.

QC of splicing for transcripts with defective BS and 3'SS acts faster, with degradation exceeding transcription from only 5 minutes post induction for Ribo1 3'SS. The transcript half-life is also longer than the 5'SS mutant, at 4.4 minutes, about the same as for the lariat of Ribo1. As 1st step has occurred in both cases an active spliceosome must have formed on these transcripts. The time taken for degradation of this RNA is probably lengthened by the need to disassemble the entire spliceosome. The more rapid checkpoint activation is presumably because the spliceosome has stalled after 1st step and the spliceosome's QC systems are primed to act. The drop in pre-mRNA after 5 minutes could be feedback from the stalled spliceosome to reduce transcription.

8.5 Recycling

The AID system is a rapid and specific system to deplete a protein with the flexibility to allow tuning the degradation profile for specific proteins (chapter 6). When applied to deplete late stage splicing factors (chapter 7), the spliceosome recycling facilities (Prp43p and the NTR complex), become overwhelmed, presumably active spliceosome levels fall and splicing stops, resulting in pre-mRNA accumulation. This happens at different times for different transcripts. Possibly there are varying capacities of spliceosomes and recycling machinery in their vicinity, but it is a global phenomenon and splicing stops for all transcripts fairly rapidly. There is some evidence that, at a very

early time points in the Prp16p depletion experiment, a checkpoint has failed to flag potential problem transcripts allowing some transcripts to splice that would otherwise have been degraded.

8.6 Degradation

The Ribo1-OFF reporter, where transcription was repressed by the addition of doxycycline, was used to study degradation (chapter 2). Curiously, transcripts that were spliced were more stable than intronless transcripts that do not require splicing. It is possible that Prp17p, or eIF4G (chapter 5) provides the link from splicing to downstream processes and influences degradation rate. The 5' and 3'SS mutant transcripts in the same study are degraded very rapidly once splicing stalled and intron retention detected.

Thio-labelling studies also produced degradation rates (chapter 4), indicating that transcript length and secondary structure strength promote stability of non-coding RNAs, and that their degradation is a multifactor process rather than a simple rate constant.

8.7 Final Words

This thesis forms a whole body of work round chapter 2, which covers aspects of RNA production, processing and degradation. Individual chapters are focussed on narrower aspects of RNA metabolism.